Design of Experiments in Stochastic Dynamics

A Thesis in Applied Mathematics submitted to the Comenius University in Bratislava in fulfilment of the requirements for the degree "philosophiae doctor" by

RNDr. Vladimír Lacko

Supervised by

Doc. Mgr. Radoslav Harman, PhD.

Co-supervised by

Prof. RNDr. Andrej Pázman, DrSc.

Defended at

Department of Applied Mathematics and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava

Bratislava, AD 2014

Abstract

The Thesis deals with the design of experiments for processes described by stochastic differential equations. The traditional approach to designing experiments is based on solving a deterministic system with subsequent contamination by a white noise, which often does not correspond to the reality. In contrast to the traditional approach, in the Thesis we assume a randomness to be an inherent element of the observed process. The adjustment of the stochastic model has a significant impact not only on the optimal allocation of observations but also on the attainable amount of information.

The main result of the Thesis is two-fold: the first is an explicit closed form of the asymptotic Fisher information matrix for linear stochastic differential equations, which can be used for computation of the ultimate efficiency of a design. On the one hand, the ultimate efficiency gives an assessment of how much a given design exhausts the ultimate information and whether an optimization of experiment is needed. On the other hand, the ultimate efficiency indicates whether the costs for performing another measurement are adequate to the gain in the amount of information. The second result of the Thesis is the discussion of the existence of optimal designs for linear stochastic differential equations, which is essential for the most basic objective of in the theory of optimal designs, and thus has theoretical and also practical meaning.

The achieved results are put into contrast with some of the recent publications, the results are partially extended to a broader class of stochastic differential equations, and we give a demonstration using the Gompertz growth model.

Keywords: Itō stochastic differential equation, exact design, product covariance structure, asymptotic Fisher information matrix, efficiency, Gompertz model.

Lacko. V: *Design of Experiments in Stochastic Dynamics*, Thesis, Comenius University

E-mail: *lackovladimir@gmail.com*
Typeset in LaTeX using the Memoir Class

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	RNDr. Vladimír Lacko
Study programme:	Applied Mathematics (Single degree study, Ph.D. III. deg., full time form)
Field of Study:	9.1.9. Applied Mathematics
Type of Thesis:	Dissertation thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Design of Experiments in Stochastic Dynamics

Aim: The aim of the thesis is the development of experimental design methods for processes described by stochastic differential equations. Since this is a new and complex topic, the primary focus of the thesis will be the analysis of the most fundamental (i.e., linear) stochastic differential equations. The existence of the optimal allocations of observations in time will be studied by means of a detailed analysis of the asymptotic properties of the information matrix. The asymptotic information matrix will also be used for the evaluation of the efficiency of finite-size sampling allocations.

Tutor:	doc. Mgr. Radoslav Harman, PhD.
Consultant:	prof. RNDr. Andrej Pázman, DrSc.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, CSc.
Assigned:	18.10.2010
Approved:	18.10.2010

prof. RNDr. Marek Fila, DrSc.
Guarantor of Study Programme

.....
Student

.....
Tutor, Consultant

Návrh experimentov v stochastickej dynamike

Dizertačná práca v odbore 9.1.9 Aplikovaná matematika odovzdaná Univerzite Komenského za účelom splnenia podmienok udelenia titulu "philosophiae doctor"

RNDr. Vladimír Lacko

Školiteľ

Doc. Mgr. Radoslav Harman, PhD.

Školiteľ-špecialista

Prof. RNDr. Andrej Pázman, DrSc.

Pracovisko

Katedra aplikovanej matematiky a štatistiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave

Bratislava, AD 2014

Abstrakt

Predložená dizertačná práca sa zaoberá navrhovaním experimentov pre procesy popísané stochastickými diferenciálnymi rovnicami. Klasický prístup k navrhovaniu experimentov spočíva v riešení deterministického systému s následnou kontamináciou bielym šumom, čo často nezodpovedá realite. Na rozdiel od klasického prístupu, v dizertačnej práci predpokladáme náhodnosť ako inherentný prvok pozorovaného procesu. Výsledná zmena stochastického modelu má zásadný vplyv nielen na optimálnu alokáciu meraní ale aj na množstvo získateľnej informácie.

Hlavné výsledky práce sú dva: Prvým je explicitný tvar asymptotickej Fisherovej informačnej matice pre lineárne stochastické diferenciálne rovnice, ktorou môžeme vyčísliť tzv. ultimátnu efektívnosť návrhu. Tá na jednej strane určí, do akej miery daný návrh vyčerpáva celkovú informáciu a či je potrebná ďalšia optimalizácia experimentu, na druhej strane indikuje, či náklady na ďalšie pozorovanie sú adekvátne vyvážené množstvom dodatočne získanej informácie. Druhým sú výsledky týkajúce sa existencie optimálnych návrhov experimentov pre lineárne stochastické diferenciálne rovnice. Opodstatnenosť tejto otázky má teoretický i praktický rozmer, nakoľko dáva zmysel základnej úlohe optimalizácie experimentu.

Uvedené výsledky sú v práci dané do kontrastu s nedávnymi publikáciami, sú čiastočne rozšírené aj na všeobecné stochastické diferenciálne rovnice, a sú demonštrované na Gompertzovom rastovom modeli.

Kľúčové slová: Itōova stochastická diferenciálna rovnica, exaktný návrh, súčinová kovariančná štruktúra, asymptotická Fisherova informačná matica, efektívnosť návrhu, Gompertzov model.

Lacko. V: *Návrh experimentov v stochastickej dynamike*, dizertačná práca, FMFI UK.

E-mail: *lackovladimir@gmail.com*
Napísané v LaTeX-u použitím štýlu Memoir

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: RNDr. Vladimír Lacko
Študijný program: aplikovaná matematika (Jednoodborové štúdium, doktorandské III. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: dizertačná
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Návrh experimentov v stochastickej dynamike / *Design of Experiments in Stochastic Dynamics*

Cieľ: Cieľom dizertačnej práce je rozvoj metód navrhovania experimentov pre procesy popísané stochastickými diferenciálnymi rovnicami. Keďže ide o novú a náročnú tému, práca bude primárne zameraná na analýzu najzákladnejších (čiže lineárnych) stochastických diferenciálnych rovníc. Detailným štúdiom asymptotických vlastností informačnej matice sa budeme snažiť zodpovedať otázku existencie optimálnej alokácie pozorovaní v čase. Pomocou odvodenej asymptotickej informačnej matice budeme vedieť ohraničiť množstvo informácie z experimentov s konečným rozsahom.

Školiteľ: doc. Mgr. Radoslav Harman, PhD.
Konzultant: prof. RNDr. Andrej Pázman, DrSc.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Spôsob sprístupnenia elektronickej verzie práce:
bez obmedzenia

Dátum zadania: 18.10.2010

Dátum schválenia: 18.10.2010 prof. RNDr. Marek Fila, DrSc.
garant študijného programu

.....
študent

.....
školiteľ, konzultant

Preface

Stochastic calculus is a field of mathematics that operates on random processes and establishes a consistent theory of integration of random processes. It is primarily focused on characterisation of the trajectories of diffusion processes, applications of which can be found in different areas of research, mainly in physics, biology/medicine and finance.

The story of the theory of stochastic calculus started in 1784, when eminent Dutch physiologist, botanist and chemist Johann Ingen-Housz [46] described the irregular movement of charcoal dust particle on the surface of alcohol. Later, in 1828, Scottish botanist Robert Brown [8] observed the same phenomenon for grains of pollen in water. Although none of the previous gentlemen provided a rigorous mathematical representation of their findings, Robert Brown already asked for it. It took more than 70 years until mathematicians started to systematically study the trajectories of continuous diffusion processes. In 1900 Louis Bachelier in his Thesis [5] for the first time mathematically defined the process of Ingen-Housz and Brown, which is in the modern literature known as "Brownian motion". Simultaneously, Albert Einstein [20] mathematically derived the same process to describe the kinetics of particles, response for which was the famous paper of Uhlenbeck and Ornstein [104]. An important milestone in further developments was the contribution of Norbert Wiener, an American mathematician and the originator of cybernetics, who proved in 1923 the existence of Brownian motion [108]; for this reason the Brownian motion is often referred to as Wiener process, and so we do in this Thesis. Since 1920's, lots of publications on stochastic dynamics have been printed out, which profiled the focus of the mathematical theory. From a number of authors, of which the most significant can be found in the monograph of Øksendal [73], a special attention is paid to the contributions of Itō, McKean, Skorokhod, Lévy, Feller, Dynkin and, of course, Kolmogorov.

Stochastic differential equations became very popular in applications, which naturally lead to an extensive progress in statistical inference, mainly in the parameter estimation. Just a brief survey on the published methods is enough for writing a monograph; see, for instance, a recent book of Iacus [45]. However, literature offers only a few publications on designing of the experiments for processes, which solve stochastic differential equations. The main reason is probably the fact that for the models given by stochastic differential equations, there is a lack for explicit solutions. In general situations we are not even able to explicitly evaluate the expectation and covariance structure of the process. Henceforth, we usually cannot formulate the problem in the standards of design of experiments.

By this Thesis I am trying to open a way for a systematic study of experiments for processes that solve stochastic differential equations from the design point of view. Although I have primarily studied linear equations, some ideas for general models are in the end available, too.

Originally, the subject of my Thesis should have been somewhat different; with the aim to contribute to the theory of exact optimal sampling designs subject to correlated observations with product covariance structures. After I made some progress by putting product

covariance structures partially into relation with some stochastic differential equations (see Chapter 5), I visited *Experiments for Processes With Time or Space Dynamics* workshop held at Isaac Newton Institute, Cambridge, UK. This conference significantly influenced the course of my further research. When my supervisor Radoslav Harman presented his latest results on product covariance structures [37] which in my view exhausted the topic enough, I decided to fully concentrate on the experimental design for stochastic dynamic systems. After all, this decision turned out to be a good one. Designing of experiments for processes governed by stochastic differential equations required a completely new way of thinking, which yielded results that were found to be key for some of my colleagues.

Here, I would like to take the opportunity to thank the people that have gone along with me. First of all and most importantly, I am grateful to my beloved wife, Andrejka, for her support, patience and boundless love. I am indebted to Doc. Radoslav Harman, with whom I co-operated already before graduation, who was a great supervisor and a good friend during difficult periods of my studies. A special thanks goes to my supervisor-specialist Prof. Andrej Pázman not only for his well-taken comments, interest and long-lasting discussions that significantly influenced my research but also for being my tutor who used his life-time experience to form my personality. I would like to acknowledge Prof. Daniel Ševčovič for intensive communication on some specific topics of this Thesis. Finally, I want to thank Doc. Viktor Witkovský for his lectures on econometrics that played a crucial role in my choice for Statistics, Prof. Dariusz Uciński for a chance to attend some of the most important scientific events in the last four years and Doc. Milan Hamala for being the first to support my interest in research in my early studies.

Bratislava, August 22, 2014

Vladimír Lacko

I dedicate this Thesis to the memory of my dear friend

Doc. PaedDr. Ing. Oliver Marton, CSc.
(⋆ 1924 – † 2013)

Contents

List of symbols *iii*

Introduction *vii*

Goals of the thesis *ix*

I Preliminaries **1**

1 Elements of stochastic calculus *2*
1.1 Basic definitions . . . 2
1.2 Kolmogorov's Extension theorem . . . 3
1.3 Wiener process . . . 4
1.4 Concept of Itō integral . . . 5
1.5 Itō stochastic differential equations . . . 8
1.5.1 Itō's lemma . . . 9
1.5.2 Existence and uniqueness of solutions . . . 9
1.6 Transition density kernel and Kolmogorov's forward equation . . . 11

2 Elements of statistical theory of experiments *13*
2.1 Fisher information . . . 13
2.2 Experimental designs, information functions and efficiency . . . 15
2.3 Regression experiments . . . 17
2.4 Basics of the theory of experimental design for regression models under uncorrelated observations . . . 18

II Literature review **20**

3 Overview of the experimental design with correlated observations *21*
3.1 Aspects of experimental design for regression models with correlated observations . . . 21
3.2 Asymptotic approach using reproducing kernel Hilbert spaces . . . 23
3.2.1 Single-parameter setup . . . 23
3.2.2 Multi-parameter setup . . . 25
3.3 Alternative asymptotic theories . . . 28
3.4 The virtual noise method of Pázman and Müller . . . 30

4 Experimental design for processes described by stochastic differential equations *31*
4.1 Models with product covariance structures . . . 31
4.2 Nonstationary Ornstein-Uhlenbeck processes . . . 32
4.3 Stationary Ornstein-Uhlenbeck processes . . . 33
4.4 Brownian motions . . . 34
4.5 Compartmental models with stochastic trajectories . . . 35
4.5.1 A naive two compartmental model with stochastic trajectories . . . 35
4.5.2 A two compartmental model with positive stochastic trajectories . . . 36

III Contribution **37**

5 Experiments with nonstationary Ornstein-Uhlenbeck processes with time-dependent volatility *38*
5.1 Corresponding linear regression model . . . 39
5.2 Information matrix and optimal designs . . . 41
5.3 Example . . . 44

6 Ultimate efficiency of experimental designs for Ornstein-Uhlenbeck type processes *46*
6.1 Nonlinear regression model for observations . . . 48
6.2 Existence of locally optimal designs . . . 51
6.3 Ultimate efficiency of designs . . . 57
6.4 Ornstein-Uhlenbeck type processes . . . 58
6.5 Example: Gompertz model of tumour growth . . . 59

7 On the information contained in the trajectory of a process governed by a stochastic differential equation *62*
7.1 Limiting behaviour of the solution to the Kolmogorov's equation . . . 63
7.2 Ultimate information . . . 64

8 Relation of the results with recent publications *65*
8.1 Concerning the existence of optimal designs . . . 65
8.2 Information matrices . . . 66

Resumé *68*
Úvod . . . 68
Ciele dizertačnej práce . . . 68
Motivácia: analýza neautonómneho nestacionárneho Ornsteinovho-Uhlenbeckovho procesu . . . 69
Výsledky pre všeobecné lineárne stochastické diferenciálne rovnice . . . 71
Model a jeho vlastnosti . . . 71
Existencia lokálne optimálnych návrhov . . . 73
Ultimátna efektívnosť návrhov . . . 74
Procesy Ornsteinovho-Uhlenbeckovho typu . . . 75
Príklad aplikácie výsledkov na Gompertzovom rastovom modeli . . . 76
Zovšeobecnenie výsledkov pre asymptotickú Fisherovu informačnú maticu . . . 77

Bibliography *78*

List of symbols

$\mathbb{N}$
: The set of natural numbers, $\{1, 2, \ldots\}$

$\mathbb{R}$
: The set of real numbers

$\mathbb{R}^n$
: an n-dimensional vector space

$\mathbb{R}^{n \times m}$
: a vector space of $n \times m$-matrices

$\mathbb{S}^n$
: The set of all $n \times n$ real symmetric matrices

$\mathbb{S}^n_+$
: The set of all $n \times n$ real symmetric non-negative definite matrices

$\mathbb{S}^n_{++}$
: The set of all $n \times n$ real symmetric positive-definite matrices

$\dim$
: Dimension

$\boldsymbol{x}, \boldsymbol{y}, \ldots$
: Vectors, usually consisting of n elements

x_i, y_i
: The ith element of the vectors $\boldsymbol{x}, \boldsymbol{y}, \ldots$

$a \in \boldsymbol{x}$, $\boldsymbol{x} \setminus \boldsymbol{y}$, $\boldsymbol{x} \subseteq \boldsymbol{y}$
: Symbols with meaning analogous to the set theory: a is equal to some element of $\boldsymbol{x}$, elements of vector $\boldsymbol{x}$ that are not elements of $\boldsymbol{y}$ with order inherited from $\boldsymbol{x}$, $\boldsymbol{x}$ consists of some elements of $\boldsymbol{y}$ with order inherited from $\boldsymbol{y}$

$\boldsymbol{X}, \boldsymbol{Y}, \ldots$
: Random vectors, usually consisting of n elements

$X_i, Y_i, \ldots$
: The ith element of the random vectors $\boldsymbol{X}, \boldsymbol{Y}, \ldots$

$\boldsymbol{X} \mid \boldsymbol{Y}$
: A random vector $\boldsymbol{X}$ conditioned on the random vector $\boldsymbol{Y}$

$\|\boldsymbol{x}\|_{\mathbf{W}}$
: A norm of a vector, $\|\boldsymbol{x}\|^2_{\mathbf{W}} \equiv \boldsymbol{x}^{\mathsf{T}} \mathbf{W}^{-1} \boldsymbol{x}$

$\|\boldsymbol{x}\|$
A usual norm of a vector, $\|\boldsymbol{x}\|^2 \equiv \boldsymbol{x}^\mathsf{T}\boldsymbol{x}$

$\boldsymbol{h}(\boldsymbol{x})$
$= (h(x_1), \ldots, h(x_n))^\mathsf{T}$ for arbitrary (random or non-random) function $h(\cdot)$ and n-vector $\boldsymbol{x}$

$\mathbf{A}, \mathbf{B}, \ldots$
Matrices

$\mathbf{A}(\boldsymbol{x}^\mathsf{T}, \boldsymbol{y})$
$= \{A(x_i, y_j)\}_{ij}$ for arbitrary (random or non-random) function $A(\cdot)$ and vectors $\boldsymbol{x}$ and $\boldsymbol{y}$

$\mathscr{M}(\mathbf{A})$
the column space of the matrix $\mathbf{A}$, linear space generated by columns of $\mathbf{A}$

$\mathscr{M}(\{f_i(t) \mid i = 1, 2, \ldots\})$
linear space generated by functions $f_i(t)$, $i = 1, 2, \ldots$

$\mathbf{A}^{-1}$
The inverse matrix to $\mathbf{A}$

$\mathbf{A}^{-}$
Arbitrary g-inverse matrix to $\mathbf{A}$

$\mathrm{tr}\{\mathbf{A}\}$
The trace of a matrix $\mathbf{A}$

$\det(\mathbf{A})$
The determinant of a matrix $\mathbf{A}$

$\mathbf{A} \succeq_\mathrm{L} \mathbf{B}$
A matrix $\mathbf{A}$ Loewner dominates matrix $\mathbf{B}$, i.e. $\mathbf{A} - \mathbf{B}$ is non-negative definite

$\Pr$
Probability

$\mathcal{E}, \mathcal{E}_{\boldsymbol{X}}$
Expectation, Expectation with respect to $\boldsymbol{X}$ (used where needed to avoid any confusion)

$\mathcal{V}$
Variance, variance-covariance matrix

$\mathcal{C}$
Covariance

$\boldsymbol{\Theta}$
A parametric space, $\boldsymbol{\Theta} \subseteq \mathbb{R}^m$

$\boldsymbol{\theta}$
An unknown parameter, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$\bar{\boldsymbol{\theta}}$
True value of the parameter

$\hat{\boldsymbol{\theta}}$

A maximum likelihood estimate of parameter $\boldsymbol{\theta}$

$\boldsymbol{\theta}^*$
A guess at the true value of parameter $\boldsymbol{\theta}$

$\mathcal{X}$
A set of experimental conditions

ξ_n, ζ_n
Exact n-point sampling design, $\xi_n = (x_1, \ldots, x_n)^\mathsf{T}$, $x_i \in \mathcal{X}$

Ξ_n
A set of all feasible exact n-point designs in experiment

ξ, ζ
An approximate design

Ξ
A set of all feasible approximate designs in experiment

$\mathcal{D}$
A sampling domain, a closed continuous interval $\mathcal{D} = [T_*, T^*]$, $0 < T_* < T^*$

$\boldsymbol{\tau}, \boldsymbol{\tau}_n$
A sampling design and an n-point sampling design, respectively

$\mathfrak{C}_\mathcal{D}$
A set of sequences covering $\mathcal{D}$. A sequence $\{\boldsymbol{\tau}^{(n)}\}_n$ is covering $\mathcal{D}$ if $t_1^{(n)} = T_*$, $t_n^{(n)} = T^*$ and $\max_{2 \le i \le n}(t_i^{(n)} - t_{i-1}^{(n)}) \to 0$.

$\mathfrak{T}_n$
A set of feasible n-point sampling designs on $\mathcal{D}$, $\mathfrak{T}_n \equiv \{\boldsymbol{\tau}_n = (t_1, \ldots, t_n)^\mathsf{T} \mid T_* \le t_1 < \ldots < t_n \le T^*\}$

$\overline{\mathfrak{T}}_n$
The closure of set $\mathfrak{T}_n$, $\overline{\mathfrak{T}}_n \equiv \{\boldsymbol{\tau}_n = (t_1, \ldots, t_n)^\mathsf{T} \mid T_* \le t_1 \le \ldots \le t_n \le T^*\}$

$\mathscr{I}, \mathscr{I}(\boldsymbol{\theta}^*)$
Fisher information matrix, Fisher information matrix evaluated at $\boldsymbol{\theta}^*$

$\mathscr{I}, \mathscr{I}(\xi), \mathscr{I}(\boldsymbol{\tau})$
Fisher information matrix corresponding to approximate design ξ and Fisher information matrix corresponding to sampling design $\boldsymbol{\tau}$, respectively

$\mathscr{I}_{\boldsymbol{X}}$
Fisher information carried by random vector $\boldsymbol{X}$

$\mathcal{M}_\Xi$
A set of all information matrices in experiment under approximate designs, $\mathcal{M}_\Xi \equiv \{\mathscr{I}(\xi), \xi \in \Xi\}$

$\mathcal{M}_{\mathfrak{T}_n}$
A set of all information matrices in experiment under sampling designs, $\mathcal{M}_{\mathfrak{T}_n} \equiv \{\mathscr{I}(\boldsymbol{\tau}_n), \boldsymbol{\tau}_n \in \mathfrak{T}_n\}$

$\{X(t)\}_{t \ge 0}$
A continuous-time scalar stochastic process

$\{W(t)\}_{t\geq 0}$
: The Wiener process

$\{\mathcal{F}_t^W\}_{t\geq 0}$
: A filtration corresponding to the Wiener process

Introduction

Many phenomena that are subject to experimental examination are of the dynamic nature. A number of examples can be found in physics, biology, social behaviour, engineering etc., where the fundamental laws are commonly described by differential equations. Although these models give a good proxy for the reality, there are always differences between theoretical and observed values.

A usual situation is depicted in Figure 1 where the lines represent theoretical trajectories and points correspond to actual states of a system in different times and different initial values. In statistical inference and design, the differences between theoretical and observed values are considered as random unobservable errors.

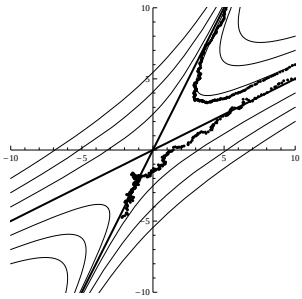


Figure 1. Example for theoretical trajectories (lines) and simulated observations (points) of a dynamic system.

There are essentially two approaches how to handle the randomness in dynamic systems. Let the law $\mathcal{L}$ describing the changes in system's state $x(t) = x_{\boldsymbol{\theta}}(t)$ at the time t depend on some parameter $\boldsymbol{\theta}$. The first approach assumes that the state $x(t)$ solves a deterministic problem exactly (hence $x(t)$ is a deterministic trajectory) and the observations are contaminated with homoscedastic and independent noise, ε_i, accredited to the measurement device, i.e., we have

$$\begin{cases} \mathcal{L}\left(\frac{\mathrm{d}^k x(t)}{\mathrm{d}t^k}, \frac{\mathrm{d}^{k-1} x(t)}{\mathrm{d}t^{k-1}}, \ldots, \frac{\mathrm{d}x(t)}{\mathrm{d}t}, x(t), t, \boldsymbol{\theta}\right) = 0, \\ X(t_i) = x(t_i) + \varepsilon_i, \ i = 1, \ldots, n. \end{cases} \tag{1}$$

Examples for this case can be found in a number of classical and novel publications; see, for instance, compartmental models in the book of Seber and Wild [97], enzyme kinetics by Michaelis and Menten[68] intensively studied by Prof. Atkinson and others, or Li's contribution [64] to the theory of D-optimal designs for Gompertz growth law [36].

In the second approach we integrate the "noise" directly into the dynamic law by considering

$$\mathcal{L}\left(\frac{\mathrm{d}^k X(t)}{\mathrm{d}t^k}, \frac{\mathrm{d}^{k-1} X(t)}{\mathrm{d}t^{k-1}}, \ldots, \frac{\mathrm{d}X(t)}{\mathrm{d}t}, X(t), t, \boldsymbol{\theta}, \text{“noise”}\right) = 0. \tag{2}$$

Of course, such notation is informal; a more rigorous formulations are given in Chapter 1 on stochastic calculus.

The main difference between the two proposals (1) and (2) lies in the interpretation of the randomness. The second formulation generates a stochastic process with correlated observations, while model (2.1) supresses the intrinsic randomness, which is characteristic for processes in the nature. Indeed, we may argue that correlation can be incorporated also in the model (1), but, unlike in formulation (2), covariance structure would be artificial.

The presented thesis is focused on the latter approach, that is, we assume the effect of the randomness to be a part of the dynamic law. The theory for such models has not been systematically built and we can find only few publications concerning such set up, where the models are rather very specified than general; see Chapter 4. Obviously, a general setup (2) would lead to a very broad class of problems, of which investigation would take one human life. Therefore, we narrow our attention to linear stochastic differential equations, but we provide also some general results.

A reasonable argument for studying linear problems is that every theory is, at the beginning, built on the most simple (linear) formulations, cf. the theory of dynamic systems, the theory of experimental design, etc. Secondly, the results for linear problems might give some indications for further extension of the results to general formulations.

Organisation of the Thesis

The Thesis is divided into three parts.

Because of the interdisciplinarity of the Thesis, Part I gives a summary of the elemental knowledge necessary for the understanding of the results. Chapter 1 provides a survey on elements of stochastic calculus, which is crucial for the formulation of the underlying models. The focal point of the 2nd chapter is the statistical background, where we recap various topics concerning the Fisher information matrix, terminology of experimental design, regression experiments and results on optimal designs under uncorrelated errors.

Part II investigates fundamental and most recent results in the theory of optimal design of experiments with correlated errors. Here, in Chapter 3 we give an outline of the main differences between the models with and without correlation which is followed by discussions on asymptotic theories and computational aspects. Chapter 4 is focused on design problems for correlated observations too, but its content is more specialised to processes described by stochastic differential equations.

The aim of Parts I and II is to give the reader rather a comprehensible insight into the subject of the Thesis than a stream of mathematically rigorous formulations.

The last Part III provides the candidate's contribution to the theory of experimental design for stochastic differential equations in a chronologic order. Chapter 5 exhibits an analysis of designs and information for the non-autonomous non-stationary Ornstein-Uhlenbeck process, which is a motivation for the subsequent chapters. In Chapter 6 we present the main results for linear stochastic differential equations. This includes evaluation of the ultimate efficiency of designs and a study on existence of optimal designs. Applications are demonstrated on the Gompertz growth law. It turns out that the ultimate information contained in the trajectories of processes described by general stochastic differential equations can be to a certain extent written in an explicit form as well, which is the topic of Chapter 7. The relations between the outcomes of Part III and recent publications are reviewed in Chapter 8.

Goals of the thesis

The subject of the Thesis related to design of experiments for processes governed by stochastic differential equations. The main aim of the Thesis is to ...

- **... study the existence of optimal sampling designs for linear stochastic differential equations**: A proof of existence of optimal sampling design is a fundamental problem, since it gives a rationale to use of the optimisation techniques and the objective of optimal design itself. A usual way for verification of the existence of optimal sampling designs is imposing of conditions which imply the continuity of the Fisher information matrix at the boundary designs $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$. For instance, Sacks and Ylvisaker [92, 93] formulated the continuity of the Fisher information matrix through the continuity of the reproducing kernel Hilbert spaces. In the majority of publications, e.g. [16, 53, 39, 40, 37] and others, where the optimal designs were proved to exist (mostly by their derivation), the authors usually assume parameters that are only present in the expectation of the process and those that are exclusively in the covariance structure. On the other hand, instances in which we share parameters in the expectation and covariance structure exhibit complications, cf. [110].

- **... derive the form of the asymptotic information matrix for linear stochastic differential equations**: Computation of optimal exact sampling designs in models with correlated observations is a non-convex problem, and explicit solutions are available only for few models, see, e.g., [16, 53, 39, 40, 37]. The other authors developed numerical procedures for identification of optimal allocation [81] or they tried to circumvent these difficulties by building alternative theories based on asymptotic arguments (in the sense of the number of design points, [92, 93, 94, 106, 107, 24, 111, 17]). In the presented Thesis we try to circumvent the most basic objective of the optimal design problem by focusing on the computation of the so-called "ultimate" efficiency [79, 37], where we compare the collected information for an exact design with the information contained in the trajectory of the process, which has a practical application [92].

- **... extend the results to the class of general stochastic differential equations**: The analysis of linear problems is always a fundamental step in order to analyse general nonlinear problems. It is of utmost interest to search for possibilities of extension of the results to general stochastic differential equations or at least to adapt developed techniques.

Part I

Preliminaries

Chapter **1**

Elements of stochastic calculus

The aim of this chapter is to give an insight into the basic concept of Itō stochastic calculus, whose language we use to formulate the underlying models later in this thesis. The most of the presented results are well known, and we refer the reader to the classical monographs by Øksendal [73], Gardiner [35] or Arnold [4].

1.1 Basic definitions

Definition 1. *A stochastic process is a collection of t-parametrised random variables $\{X(t,\omega)\}_{t\in T}$ defined on some probability space $(\Omega, \mathcal{F}, \Pr)$ with values in $\mathbb{R}$.*

For the continuous-time stochastic process, the set T is usually an interval $[a,b]$ or a half-line $[0,\infty)$. In the latter instance, instead of $\{X(t)\}_{t\in T}$, we use the notation $\{X(t)\}_{t\geq 0}$. Although Definition 1 only gives a notation for scalar-valued processes, we are not limited in the number of dimensions. To understand the concept of Itō integral outlined in this chapter, the one-dimensional case is sufficient. Nevertheless, in some important cases we introduce also multivariate counterparts.

Definition 1 designates stochastic processes as collections of random variables, where, for a given $t \in T$, we have the mapping $X(t,\cdot) : \Omega \mapsto \mathbb{R}$. On the other hand, for a fixed $\omega \in \Omega$, the mapping $X(\cdot,\omega) : T \mapsto \mathbb{R}$ is a trajectory or path of the stochastic process.

Definition 2. *A stochastic process $\{Y(t,\omega)\}_{t\in T}$ is a version of the stochastic process $\{X(t,\omega)\}_{t\in T}$ on $(\Omega, \mathcal{F}, \Pr)$ if for all $t \in T$, $\Pr[\{\omega \in \Omega \mid Y(t,\omega) = X(t,\omega)\}] = 1$.*

Definition 3. *A filtration on the measurable space $(\Omega, \mathcal{F})$ is a sequence $\{\mathcal{F}_t\}_{t\in T}$ of σ-algebras such that $\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2} \subseteq \mathcal{F}$ for all $t_1, t_2 \in T$, $t_1 \leq t_2$. The quadruplet $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in T}, \Pr)$ is called a filtered probability space.*

Filtration represents an information flow (not in a statistical sense), or the history of the process. It determines, which event we are able to discriminate in terms of measurable cylinders in classical probability theory; see [56].

Definition 4. *Let $\{\mathcal{F}_t\}_{t\in T}$ be a filtration on $(\Omega, \mathcal{F})$ and let $\{X(t,\omega)\}_{t\in T}$ be a stochastic process on $(\Omega, \mathcal{F}, \Pr)$. The process $\{X(t,\omega)\}_{t\in T}$ is said to be $\mathcal{F}_t$-adapted if for each $t \in T$, the mapping $\omega \mapsto X(t,\omega)$ is $\mathcal{F}_t$-measurable.*

For the sake of simplicity in the rest of this chapter we omit a notation of the dependence of a stochastic process on $\omega \in \Omega$. The only exception is Section 1.4, where we define stochastic integrals.

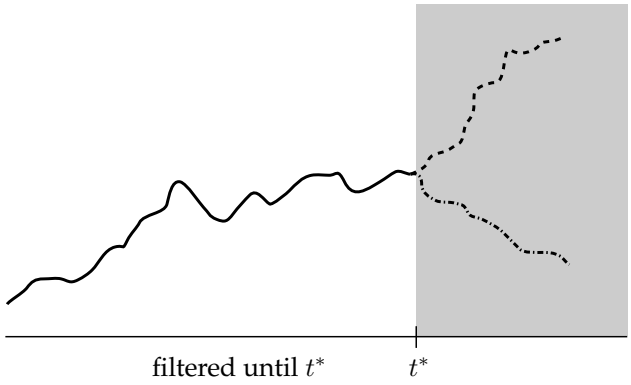


Figure 1.1. Filtration represents an information flow about events that could be distinguished. Figure depicts two different possible paths of a stochastic process. Until time t^*, where the process is filtered we cannot distinguish these paths, because they differ only at times after t^* (dashed and dash-dotted lines in the grey area). To recognize these paths we have to move further in time, when the corresponding σ-algebra is more rich.

Definition 5. *By a sampling n-point design $\boldsymbol{\tau}_n = (t_1, \ldots, t_n)^\mathsf{T}$ on the domain $\mathcal{D} \subseteq T$ we understand a sequence of n non-decreasing times from $\mathcal{D}$. We further define $\mathfrak{T}_n \equiv \{\boldsymbol{\tau}_n = (t_1, \ldots, t_n)^\mathsf{T} \mid T_* \leq t_1 < \ldots < t_n \leq T^*\}$ and $\overline{\mathfrak{T}}_n \equiv \{\boldsymbol{\tau}_n = (t_1, \ldots, t_n)^\mathsf{T} \mid T_* \leq t_1 \leq \ldots \leq t_n \leq T^*\}$ is the closure of $\mathfrak{T}_n$.*

1.2 Kolmogorov's Extension theorem

A fundamental result with impact on stochastic calculus and stochastic processes is the Kolmogorov's Extension Theorem [56]. The theorem was originally formulated for discrete sequences of random variables, but in the modern literature on stochastic differential equations we can usually find the form applicable to continuous stochastic processes; see, e.g., [73].

Proposition 1 (Kolmogorov's Extension Theorem)**.** *For all $t_1, \ldots, t_n \in T$, $n \in \mathbb{N}$, let $\nu_{t_1,\ldots,t_n}$ be probability measures on $\mathbb{R}^n$ such that*

$$\nu_{\pi(t_1),\ldots,\pi(t_n)}(A_1 \times \cdots \times A_n) = \nu_{t_1,\ldots,t_n}(A_{\pi^{-1}(1)} \times \cdots \times A_{\pi^{-1}(n)})$$

for all permutations π on $\{1, \ldots, n\}$, and

$$\nu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n) = \nu_{t_1,\ldots,t_n,t_{n+1},\ldots,t_{n+p}}(A_1 \times \cdots \times A_n \times \mathbb{R} \times \cdots \times \mathbb{R})$$

for all $p \in \mathbb{N}$, where the set on the right hand side has a total of $n + p$ factors. Then there exists a probability space $(\Omega, \mathcal{F}, \Pr)$ and a stochastic process $\{X(t)\}_{t\in T}$ on $(\Omega, \mathcal{F}, \Pr)$, $X(t) : \Omega \mapsto \mathbb{R}$, such that

$$\nu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n) = \Pr[X(t_1) \in A_1, \ldots, X(t_n) \in A_n]$$

for all $t_i \in T$, $n \in \mathbb{N}$ and all Borel sets A_i.

For a given stochastic process $\{X(t)\}_{t\in T}$, some design $\boldsymbol{\tau} = (t_1, \ldots, t_n)^\mathsf{T}$ on T and measurable sets $A_1, \ldots, A_n$, we can theoretically determine the probability that $X(t_1) \in A_1, \ldots, X(t_n) \in A_n$. The Extension Theorem solves an opposite problem: if we are given a probability space and the consistency conditions are satisfied, then there exists a random process following this probability law.

1.3 Wiener process

Definition 6. *A one-dimensional Wiener process $\{W(t)\}_{t\geq 0}$ on the probability space $(\Omega, \mathcal{F}, \Pr)$ is a continuous-time stochastic process characterized by the following three facts:*

i) $\Pr[W(0) = 0] = 1$,

ii) for any design $\boldsymbol{\tau} = (t_1, \ldots, t_n)^\mathsf{T}$ on $[0, \infty)$, the increments $W(t_n) - W(t_{n-1}), \ldots, W(t_2) - W(t_1)$ are mutually independent,

iii) for any positive constants t and Δ, the distribution of $W(t + \Delta) - W(t)$ is $\mathcal{N}(0, \Delta)$.

An n-dimensional Wiener process $\{\boldsymbol{W}(t)\}_{t\geq 0}$ is a process, for which the components correspond to n independent one-dimensional Wiener processes.

The first condition of Definition 6 is only technical, and such process is sometimes called "mathematical Brownian motion" [96, p39]. Conditions ii) and iii) give a finite-dimensional distribution of $\{W(t)\}_{t\geq 0}$, which is consistent in the sense of the Kolmogorov's Extension Theorem. Hence, there exists a probability space, where $\{W(t)\}_{t\geq 0}$ can be constructed. Notice that the Wiener process is not defined uniquely.

The Wiener process has an exceptional position amongst stochastic processes, not only for its special properties, but also from the historical aspect. The second condition in Definition 6 could be underset by interactions of particles in thermodynamics. The last property on Gaussian distribution of increments can be explained at least in two ways.

The first interpretation comes from the Einstein's work in thermodynamics [20], which does not study an individual particle but a whole cloud of particles. There he states a basic diffusion equation for the density of particles, which is solved by the Gaussian kernel. A very simple but detailed explanation of Einstein's idea based on the fact that the Wiener process is a continuous limit of a random walk can be found also in the monograph of Kwok [57]. If $X_1, \ldots, X_n$ are independent taking values only ± 1 with equal probability and if we set $S_n = X_1 + \cdots + X_n$ the partial sum, then from Central Limit Theorem for $n \to \infty$ (see, e.g., [45, p20])

$$\Pr\left[\frac{S_{[nt]}}{\sqrt{n}} < x\right] \to \Pr[W(t) < x],$$

where $[nt]$ stands for integer part of nt.

Other view, again motivated by physics, stays in a multi-dimensional view. More precisely, Gaussian distribution is, according to the Herschel-Maxwell theorem [9], the only spherically invariant distribution with independent coordinates, which is a rational requirement: if we have a uniformly distributed direction and the coordinates are mutually independent, then the Gaussian distribution is the only solution.

An important question concerning the Wiener process is about the continuity of its trajectory. The answer is proposed by Kolmogorov's Continuity Theorem [101].

Proposition 2 (Kolmogorov's Continuity Theorem)**.** *Let a process $\{X(t)\}_{t\geq 0}$ satisfy the following condition: For all $S > 0$ there exist positive constants α, β, D such that*

$$\mathcal{E}[|X(t) - X(s)|^\alpha] \leq D|t - s|^{1+\beta},\ 0 \leq s, t \leq S.$$

Then there exists a continuous version of $\{X(t)\}_{t\geq 0}$.

For the n-dimensional Wiener process it can be proved [73, p14] that

$$\mathcal{E}[\|\boldsymbol{W}(t) - \boldsymbol{W}(s)\|^4] = n(n+1)|t-s|^2.$$

The reader can find more information on the continuity in the book of Durrett [18].

Proposition 3. *For the Wiener process we have that*

i) $\mathcal{E}[W(s)W(t)] = \min\{s,t\},\ 0 \leq s,t,$

ii) there exists a continuous version,

iii) the paths are Pr*-almost surely nowhere differentiable.*

An argument for the third part of the previous proposition can be found in quadratic variation of functions.

Definition 7. *Let a sequence of designs* $\{\boldsymbol{\tau}^{(n)}\}_n \in \mathfrak{C}_{\mathcal{D}}$ *be covering* $\mathcal{D}$*. The quadratic variation of a function* $f(t)$ *on* $\mathcal{D}$ *is*

$$[f]_{[a,b]} = \lim_{n\to\infty} \sum_{k=2}^{n} [f(t_k^{(n)}) - f(t_{k-1}^{(n)})]^2 \tag{1.1}$$

It straightforward that for any differentiable function the quadratic variation is identically equal to zero.

On the other hand, for the Wiener process starting from 0 we can show [73, p19] that the quadratic variation of $\{W(t)\}_{t\geq 0}$ on $[0,s]$ is Pr-almost surely s (i.e., the sum in (1.1) converges with respect to L^2 space). Consequently, if we take into account the definition of quadratic variation (1.1), for the Wiener process we can informally write that

$$\int_a^b [\mathrm{d}W(t)]^2 = b - a = \int_a^b \mathrm{d}t,$$

or in even more informal notation

$$[\mathrm{d}W(t)]^2 \approx \mathrm{d}t. \tag{1.2}$$

That is, if the norm of the partition induced by the sequence of designs $\{\boldsymbol{\tau}^{(n)}\}_n$ tends to zero, then the behaviour of $[\mathrm{d}W(t)]^2$ becomes non-random. An illustrative numerical example is depicted in Figure 1.2.

1.4 Concept of Itō integral

The integration theory is one of the most important concepts in mathematics, and a number of variants have been developed. The ground for this thesis is a consistent integration of stochastic processes with respect to stochastic processes of the form

$$\int_a^b f(t,\omega)\mathrm{d}W(t,\omega). \tag{1.3}$$

Note that in the standard (Riemann) integration theory if a function $G(t)$ is differentiable, then

$$\int_a^b f(t)\mathrm{d}G(t) = \int_a^b f(t)\frac{\mathrm{d}G}{\mathrm{d}t}(t)\mathrm{d}t.$$

However, for a fixed $\omega \in \Omega$ the function $t \mapsto W(t,\omega)$ is not differentiable, henceforth we cannot exploit the standard integration theory. For a question what "$\mathrm{d}W(t,\omega)$" is, we have a formal answer:

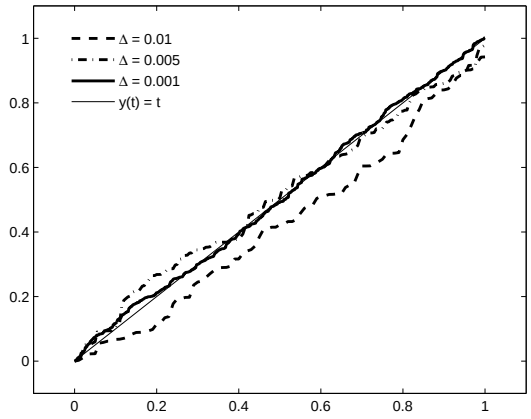


Figure 1.2. A simulation of quadratic variation process for the Wiener process by discretization (cf. (1.1)) for different time steps. Evidently, if the difference $\Delta = t_k - t_{k-1}$ is becoming smaller, the quadratic variation becomes "less stochastic" and converges to line $y(t) = t$.

Definition 8. *Let $\{W(t,\omega)\}_{t\geq 0}$ be the Wiener process. Then*

$$\int_a^b \mathrm{d}W(t,\omega) = W(b,\omega) - W(a,\omega).$$

The way of integrating of stochastic processes with respect to stochastic processes given in (1.3) answers the concept of Itō integral based on Definition 8. As usual, the Itō integration theory is built up on elementary processes (functions). Clearly, for a constant process $\{f(t,\omega)\}_{t\geq 0} = \{c\}_{t\geq 0}$ we have

$$\int_a^b f(t,\omega)\mathrm{d}W(t,\omega) = \int_a^b c\mathrm{d}W(t,\omega) = c\int_a^b \mathrm{d}W(t,\omega) = c[W(b,\omega) - W(a,\omega)].$$

Note that $c\int_a^b \mathrm{d}W(t,\omega) \sim \mathcal{N}(0, \int_a^b f^2(t,\omega)\mathrm{d}t)$. For an elementary process

$$\{f_n(t,\omega)\}_{t\geq 0} = \left\{\sum_{i=0}^{n-1} f(t_i,\omega) I_{[t_i,t_{i+1})}(t)\right\}_{t\geq 0},$$

where I_A is the indicator function, $t_0 = a$, $t_n = b$ and $t_i < t_{i+1}$ for all i, we can define

$$\int_a^b f_n(t,\omega)\mathrm{d}W(t,\omega) = \sum_{i=0}^{n-1} f(t_i,\omega)[W(t_{i+1},\omega) - W(t_i,\omega)]. \tag{1.4}$$

Again, we find out that $\int_a^b f_n(t,\omega)\mathrm{d}W(t,\omega) \sim \mathcal{N}(0, \int_a^b f^2(t,\omega)\mathrm{d}t)$. It is, therefore, natural to define the following class of processes:

Definition 9. *A process $\{f(t,\omega)\}_{t\geq 0}$ is Itō integrable on $[a,b]$ if*

i) $\{f(t,\omega)\}_{t\geq 0}$ is $\mathcal{F}_t^W$-adapted,

ii) the function $(t,\omega) \mapsto f(t,\omega)$ is Borel-measurable in t and $\mathcal{F}$-measurable in ω,

iii) $\mathcal{E}[\int_a^b f^2(t,\omega)\mathrm{d}t] < \infty$.

Then we can define an Itō integral for a process $\{f(t,\omega)\}_{t\geq 0}$ as a limit of integrals for elementary functions in (1.4):

Definition 10. *Let* $\{\{f_n(t,\omega)\}_{t\geq 0}\}_n$ *be a sequence of elementary processes such that*

$$\mathcal{E}\left[\int_a^b [f(t,\omega) - f_n(t,\omega)]^2 \mathrm{d}t\right] \to 0 \text{ as } n \to \infty. \tag{1.5}$$

(If $\{f(t,\omega)\}_{t\geq 0}$ *is Itō integrable, then* $\{\{f_n(t,\omega)\}_{t\geq 0}\}_n$ *described in* (1.5) *always exists [73].) Then we define*

$$\int_a^b f(t,\omega)\mathrm{d}W(t,\omega) = \lim_{n\to\infty} \int_a^b f_n(t,\omega)\mathrm{d}W(t,\omega),$$

where the limit is with respect to $L^2(\mathrm{Pr})$.

Proposition 4 (Itō isometry)**.** *Let* $\{f(t,\omega)\}_{t\geq 0}$ *be an Itō integrable function on* $[a,b]$. *Then*

$$\mathcal{E}\left[\int_a^b f(t,\omega)\mathrm{d}W(t,\omega)\right] = 0$$

and

$$\mathcal{E}\left[\left(\int_a^b f(t,\omega)\mathrm{d}W(t,\omega)\right)^2\right] = \mathcal{E}\left[\int_a^b f^2(t,\omega)\mathrm{d}t\right].$$

We should discuss several issues on stochastic integrals. Firstly, in the case of Itō integral the process $\{f(t,\omega)\}_{t\geq 0}$ has to be $\mathcal{F}_t^W$-adapted. As an alternative, we can take the concept of the Stratonovich integral, which, in the approximation by simple functions, evaluates $f(t,\omega)$ at $t_i^* = (t_i + t_{i+1})/2$. In contrast with a Riemann integral, Itō and Stratonovich integrals do not necessarily converge to each other [73, p35].

Secondly, the Itō integral preserves the properties of integral we are used to from standard integration theories.

Proposition 5. *Let* $\{f(t,\omega)\}_{t\geq 0}$ *and* $\{g(t,\omega)\}_{t\geq 0}$ *be Itō integrable on* $[a,b]$, $a < c < b$, *and let* $\alpha, \beta \in \mathbb{R}$. *Then*

i) $\int_a^b f(t,\omega)\mathrm{d}W(t,\omega) = \int_a^c f(t,\omega)\mathrm{d}W(t,\omega) + \int_c^b f(t,\omega)\mathrm{d}W(t,\omega)$ *for* Pr*-almost all* ω,

ii) $\int_a^b [\alpha f(t,\omega) + \beta g(t,\omega)]\mathrm{d}W(t,\omega) = \alpha \int_a^b f(t,\omega)\mathrm{d}W(t,\omega) + \beta \int_a^b g(t,\omega)\mathrm{d}W(t,\omega)$ *for* Pr*-almost all* ω,

iii) $\int_a^b f(t,\omega)\mathrm{d}W(t,\omega)$ *is a* $\mathcal{F}_b^W$*-measurable random variable,*

iv) if $f(t,\omega) = f(t)$, *then* $\int_0^t f(\nu)\mathrm{d}W(\nu,\omega) = f(t)W(t,\omega) + \int_0^t W(t,\omega)\mathrm{d}f(t)$ *for* Pr*-almost all* ω *(integration by parts).*

For more details on the properties of stochastic integrals we refer the reader to Durrett [18] and Øksendal [73].

1.5 Itō stochastic differential equations

Differential equations are often used to describe dynamics around us. In the deterministic theory of dynamic systems we can often find models characterized by equations of the form

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \dot{X}(t) = f[t, X(t)],$$

with $X(0)$ given. Nevertheless, in the real world this dynamics is usually perturbed by random interactions, the size of which might depend on the current state $X(t)$, time t or any other factor. That is, in an informal notation we can write

$$\dot{X}(t) = f[t, X(t)] + \sigma[t, X(t)]N(t), \tag{1.6}$$

where $N(t)$ is a "noise". A reasonable properties that the process $\{N(t)\}_{t\geq 0}$ should possess (at least approximately) are [73, p21]:

P1) for $t_1, t_2 \geq 0$, $t_1 \neq t_2$, the random variables $N(t_1)$ and $N(t_2)$ are independent,

P2) $\{N(t)\}_{t\geq 0}$ is a stationary process with $\mathcal{E}[N(t)] = 0$ for all t.

Any process $\{N(t)\}_{t\geq 0}$ with properties P1 and P2 is not continuous, and if it has a unit variance, the function $(t, \omega) \mapsto N(t, \omega)$ is not even measurable.

By studying a discretized version of the equation (1.6)

$$X(t_{k+1}) - X(t_k) = f[t_k, X(t_k)]\Delta t_k + \sigma[t_k, X(t_k)]N(t_k)\Delta t_k,$$

where $\Delta t_k = t_{k+1} - t_k$. Let $N(t_k)\Delta t_k = \Delta V(t_k) = V(t_{k+1}) - V(t_k)$, we get that the assumptions P1 and P2 on $\{N(t)\}_{t\geq 0}$ imply that $V(t)$ should have zero mean stationary and independent increments. Nevertheless, the only continuous process with zero mean, stationary and independent increments is the Wiener process [54].

If we sum up the discretized version of (1.6) and take the limit in the sense of Definition 10, we obtain

$$X(t) = X(0) + \int_0^t f[\nu, X(\nu)]\mathrm{d}\nu + \int_0^t \sigma[\nu, X(\nu)]\mathrm{d}W(\nu), \tag{1.7}$$

or, in the differential form

$$\mathrm{d}X(t) = f[t, X(t)]\mathrm{d}t + \sigma[t, X(t)]\mathrm{d}W(t). \tag{1.8}$$

The measurable function $f(t, x)$ is usually called a "drift", the measurable function $\sigma(t, x)$ is a "volatility" or "diffusion coefficient", and $X(0)$ is the initial value. Analogously, we can define a multivariate Itō stochastic differential equation

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{f}[t, \boldsymbol{X}(t)]\mathrm{d}t + \boldsymbol{\sigma}[t, \boldsymbol{X}(t)]\mathrm{d}\boldsymbol{W}(t).$$

Here, $\boldsymbol{f} : [0, \infty] \times \mathbb{R}^n \mapsto \mathbb{R}^n$, $\boldsymbol{\sigma} : [0, \infty] \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$ and $\{\boldsymbol{W}(t)\}_{t\geq 0}$ is the n-dimensional Wiener process.

In the following we define a class of Itō processes [73, p44].

Definition 11. *Let $\{W(t, \omega)\}_{t\geq 0}$ be the Wiener process on $(\Omega, \mathcal{F}, \Pr)$, and let $\{u(t, \omega)\}_{t\geq 0}$ and $\{v(t, \omega)\}_{t\geq 0}$ be $\mathcal{F}_t^W$-adapted processes satisfying*

$$\Pr\left[\int_0^t |u(\nu, \omega)|\mathrm{d}\nu < \infty \ \forall_{t\geq 0}\right] = 1, \ \Pr\left[\int_0^t v^2(\nu, \omega)\mathrm{d}\nu < \infty \ \forall_{t\geq 0}\right] = 1.$$

Then the one-dimensional Itō process is a stochastic process $\{X(t)\}_{t\geq 0}$ *on* $(\Omega, \mathcal{F}, \Pr)$ *of the form*

$$X(t) = X(0) + \int_0^t u(\nu,\omega)\mathrm{d}\nu + \int_0^t v(\nu,\omega)\mathrm{d}W(\nu,\omega).$$

Similarly to the Wiener process, we can calculate the quadratic variation also for Itō processes [45, p37]:

Proposition 6. *Let* $\{X(t)\}_{t\geq 0}$ *be a one-dimensional Itō process. Then the quadratic variation of* $\{X(t)\}_{t\geq 0}$ *on* $\mathcal{D} = [a, b]$, $0 < a < b$, *is*

$$[X]_{\mathcal{D}} = \int_a^b v^2(t,\omega)\mathrm{d}t.$$

1.5.1 Itō's lemma

Itō's lemma [47, Thm6], which is an important result of stochastic calculus, was already sketched by formula (1.2). That is, in contrast with differentiable functions, in total differentials of Itō processes we have to take into account also second order terms.

Proposition 7 (Itō's lemma)**.** *Let* $\{X(t)\}_{t\geq 0}$ *be an Itō process described by a stochastic differential equation of the form*

$$\mathrm{d}X(t) = f[t, X(t)]\mathrm{d}t + \sigma[t, X(t)]\mathrm{d}W(t),$$

and $g(t, x)$ *be twice continuously differentiable on* $[0,\infty)\times\mathbb{R}$. *Then the process* $\{Y(t)\}_{t\geq 0}$ *given by*

$$Y(t) = g[t, X(t)]$$

is an Itō process as well, and

$$\begin{aligned}\mathrm{d}Y(t) &= \frac{\partial g}{\partial t}\Big|_{[t,X(t)]}\mathrm{d}t + \frac{\partial g}{\partial x}\Big|_{[t,X(t)]}\mathrm{d}X(t) + \frac{1}{2}\sigma^2[t,X(t)]\frac{\partial^2 g}{\partial x^2}\Big|_{[t,X(t)]}\mathrm{d}t\\ &= \left(\frac{\partial g}{\partial t} + f[t,X(t)]\frac{\partial g}{\partial x} + \frac{1}{2}\sigma^2[t,X(t)]\frac{\partial^2 g}{\partial x^2}\right)\Big|_{[t,X(t)]}\mathrm{d}t + \sigma[t,X(t)]\mathrm{d}W(t).\end{aligned}$$

A multivariate counterpart is at hand: we have a process described by an equation $\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{f}[t,\boldsymbol{X}(t)]\mathrm{d}t + \boldsymbol{\sigma}[t,\boldsymbol{X}(t)]\mathrm{d}\boldsymbol{W}(t)$ and $Y(t) = g[t,\boldsymbol{X}(t)]$, where the function $g(t,\boldsymbol{x})$ is scalar-valued and twice continuously differentiable on $[0,\infty)\times\mathbb{R}^n$. Then

$$\mathrm{d}Y(t) = \frac{\partial g}{\partial t}\Big|_{[t,\boldsymbol{X}(t)]}\mathrm{d}t + \frac{\partial g}{\partial \boldsymbol{x}}\Big|_{[t,\boldsymbol{X}(t)]}\mathrm{d}\boldsymbol{X}(t) + \frac{1}{2}[\mathrm{d}\boldsymbol{X}(t)]^{\mathsf{T}}\frac{\partial g}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathsf{T}}}\Big|_{[t,\boldsymbol{X}(t)]}\mathrm{d}\boldsymbol{X}(t), \tag{1.9}$$

where we set $\mathrm{d}W_i(t)\mathrm{d}t = (\mathrm{d}t)^2 = 0$ and $\mathrm{d}W_i(t)\mathrm{d}W_j(t) = \delta_{ij}\mathrm{d}t$, and δ_{ij} is the Kronecker delta. If g is a vector function, that is, we have $\boldsymbol{Y}(t) = \boldsymbol{g}[t,\boldsymbol{X}(t)]$, then we compute (1.9) for each component g_i.

1.5.2 Existence and uniqueness of solutions

A solution to a stochastic differential equation is a stochastic process. Similarly to the theory of deterministic dynamic systems, we should ask whether any solution exists and if it is unique or not. In contrast with deterministic systems, for stochastic processes, we formulate the existence and uniqueness in terms of probability. We distinguish two types of solutions: weak solutions and strong solutions.

Definition 12. *A weak solution to a stochastic differential equation*

$$\begin{aligned} \mathrm{d}X(t) &= f[t, X(t)]\mathrm{d}t + \sigma[t, X(t)]\mathrm{d}W(t), \\ X(0) &\sim Z \end{aligned}$$

on a measurable filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t^W\}_{t\geq 0}, \mathrm{Pr})$*, where* $\{W(t)\}_{t\geq 0}$ *is the Wiener process and* Z *is a distribution independent from* $\{W(t)\}_{t\geq 0}$*, is a pair* $\{(X(t), W(t))\}_{t\geq 0}$ *of processes such that* $\{X(t)\}_{t\geq 0}$ *is* $\mathcal{F}_t^W$*-adapted and*

$$X(t) = X(0) + \int_0^t f[\nu, X(\nu)]\mathrm{d}\nu + \int_0^t \sigma[\nu, X(\nu)]\mathrm{d}W(\nu)$$

Pr*-almost surely for all* $t \geq 0$*.*

Definition 13. *A strong solution to a stochastic differential equation*

$$\begin{aligned} \mathrm{d}X(t) &= f[t, X(t)]\mathrm{d}t + \sigma[t, X(t)]\mathrm{d}W(t), \\ X(0) &\sim Z \end{aligned}$$

on a measurable filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t^W\}_{t\geq 0}, \mathrm{Pr})$*, with* $\{W(t)\}_{t\geq 0}$ *being a fixed Wiener process and* Z *being a distribution independent from* $\{W(t)\}_{t\geq 0}$*, is an* $\mathcal{F}_t^W$*-adapted process* $\{X(t)\}_{t\geq 0}$ *such that*

$$X(t) = X(0) + \int_0^t f[\nu, X(\nu)]\mathrm{d}\nu + \int_0^t \sigma[\nu, X(\nu)]\mathrm{d}W(\nu)$$

Pr*-almost surely for all* $t \geq 0$*.*

The main difference between the definitions of weak and strong solutions consists in the selection of the underlying Wiener process. If we seek for a weak solution, we give coefficients $f(t, x)$ and $\sigma(t, x)$ and ask for a pair $\{(X(t), W(t))\}_{t\geq 0}$, which solves the corresponding stochastic differential equation. That is, we simultaneously seek for $\{X(t)\}_{t\geq 0}$ and a version of $\{W(t)\}_{t\geq 0}$, and the weak existence and uniqueness are meant to be with respect to the probability law. On the other hand, if the version of the process $\{W(t)\}_{t\geq 0}$ is fixed, then the obtained solution $\{X(t)\}_{t\geq 0}$ is strong and the existence and uniqueness are meant to be pathwise.

Proposition 8 (Existence and uniqueness of the strong solution)**.** *Let* $\alpha, \beta \in \mathbb{R}$ *be some constants, the coefficients* $f(t, x)$ *and* $\sigma(t, x)$ *be measurable functions from* $[0, T] \times \mathbb{R}$ *to* $\mathbb{R}$*, and let* Z *be a random variable such that*

i) (at most linear growth of coefficients in x*)*

$$\forall_{x\in\mathbb{R}}\forall_{t\in[0,T]}\ |b(t, x)| + |\sigma(t, x)| \leq \alpha(1 + |x|),$$

ii) (Lipschitz continuity of coefficients)

$$\forall_{x,y\in\mathbb{R}}\forall_{t\in[0,T]}\ |b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq \alpha(1 + |x|),$$

iii) Z *is independent of* σ*-algebra* $\mathcal{F}_t^W$ *and has a finite second moment.*

Then the stochastic differential equation

$$\begin{aligned} \mathrm{d}X(t) &= f[t, X(t)]\mathrm{d}t + \sigma[t, X(t)]\mathrm{d}W(t),\ 0 \le t \le T, \\ X(0) &= Z, \end{aligned}$$

has the unique $\mathcal{F}_t^W$-adapted solution $\{X(t)\}_{t\ge 0}$ generated by $W(t)$ and Z with

$$\mathcal{E}\left[\int_0^T |X(t)|^2 \mathrm{d}t\right] < \infty.$$

The multivariate case is very similar, with the Frobenius norm used instead of the absolute value $|\cdot|$.

The strong solution is, of course, also a weak solution, but the inverse implication is not true in a general case. For the existence of a weak solution, the assumption i) of Proposition 8 is sufficient; see [98, 99, 43].

If we say that a process $\{X(t)\}_{t\ge 0}$ is described by some stochastic differential equation, we understand that $\{X(t)\}_{t\ge 0}$ is its weak solution.

1.6 Transition density kernel and Kolmogorov's forward equation

In practical applications concerning stochastic differential equations, especially in simulation, inference and experimental design, the knowledge of a transition density kernel is useful. Let $\{X(t)\}_{t\ge 0}$ be a stochastic process, under the transition density kernel we understand

$$p(x, s \mid y, t) = \frac{\mathrm{d}}{\mathrm{d}x} \Pr[X(t+s) < x \mid X(t) = y].$$

For the sake of simplicity we will discuss only the univariate case. Analogues for a vector-valued Itō processes can be found in references cited below.

A fundamental result in this field is the Kolmogorov's forward equation [55, p458]:

Proposition 9. *Let process $\{X(t)\}_{t\ge 0}$ be a solution to a stochastic differential equation of the form*

$$\mathrm{d}X(t) = f[t, X(t)]\mathrm{d}t + \sigma[t, X(t)]\mathrm{d}W(t).$$

Then the transition density kernel $p(x, s \mid y, t)$ solves the initial problem

$$\begin{cases} \dfrac{\partial p}{\partial s} + \dfrac{\partial}{\partial x}[f(t+s, x)p] &= \dfrac{1}{2}\dfrac{\partial^2}{\partial x^2}[\sigma^2(t+s, x)p], \\ p(x, 0 \mid y, t) &= \delta(x - y), \end{cases} \tag{1.10}$$

where $\delta(\cdot)$ is the Dirac delta function.

Kolmogorov's forward equation is often referred to as Fokker-Planck equation, named after its inventors Adriaan D. Fokker, who derived the law only for stationary processes [30], and Max Planck, who formulated the equation (1.10) in his publication [82]. For a detailed historical discussion we recommend the paper of Ebeling et al. [19].

The existence and uniqueness of solutions to the initial problem (1.10) was extensively studied (see, for instance, [91, 32]), but is beyond the scope of the presented thesis; henceforth, we do not provide a detailed discussion of this topic. Besides the original publications [82, 55], the reader can find the derivation of this law also in [96, 73] and many other publications. Note also that for the Wiener process, where the diffusion coefficient σ is a

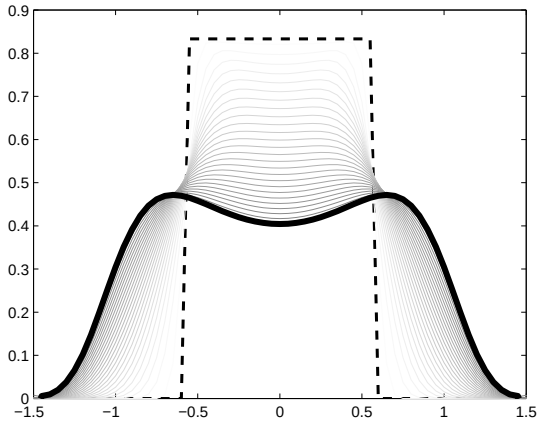


Figure 1.3. An evolution of a solution to the Kolmogorov's forward equation for the model $\mathrm{d}X(t) = 5[1 - X(t)]X(t)[1 + X(t)]\mathrm{d}t + \mathrm{d}W(t)$ with initial distribution uniform on $[-0.5, 0.5]$ (dashed line).

constant and the drift coefficient f is zero, the Kolmogorov's forward equation reduces to the standard diffusion equation derived by Einstein in his 1905 paper [20].

A matter of concern is the fact that, with exception of very special stochastic differential equations, we are usually not able to find a closed form solution for Kolmogorov's forward equation. For this reason scientists developed a variety of approximate analytical and numerical methods not only for a density evaluation itself, but also for process simulation and parameter estimation under discrete observations.

Figure 1.3 depicts an evolution of the transition density kernel for the equation $\mathrm{d}X(t) = 5[1 - X(t)]X(t)[1 + X(t)]\mathrm{d}t + \mathrm{d}W(t)$ with the initial distribution uniform on $[-0.5, 0.5]$. The shape of the density is usually connected with the properties of the deterministic counterpart of the model. For instance, in this case the deterministic model $\mathrm{d}X(t) = 5[1 - X(t)]X(t)[1 + X(t)]\mathrm{d}t$ has three stationary points $-1, 0, 1$. The stationary points $-1, 1$ are attractors and 0 is a repellor; consequently, the probability mass represented by the transition kernel is bimodal and more concentrated around the values $-1, 1$.

The transition density kernel also describes the limiting role of coefficients of stochastic differential equation. The following can be shown [55, 96]:

Proposition 10. *Let $p(x, s \mid y, t)$ solve the initial problem* (1.10)*. Then*

$$\begin{aligned}
\lim_{s \to 0} \frac{1}{s} \int_{\mathbb{R}} (y - x) p(y, s \mid x, t) \mathrm{d}y &= f(t, x), \\
\lim_{s \to 0} \frac{1}{s} \int_{\mathbb{R}} (y - x)^2 p(y, s \mid x, t) \mathrm{d}y &= \sigma^2(t, x), \\
\lim_{s \to 0} \frac{1}{s} \int_{\mathbb{R}} |y - x|^{2+\epsilon} p(y, s \mid x, t) \mathrm{d}y &= 0, \ \epsilon > 0.
\end{aligned}$$

Chapter **2**

Elements of statistical theory of experiments

Throughout this chapter we briefly review several important topics in the statistics and theory of experiments, which have become a "folklore" in the literature. We focus on the problems concerning point estimation, which includes a discussion on Fisher information and introduction to design of regression experiments under uncorrelated observations. The reason for the last topic is that the problems analysed in Parts II and III assume correlated observations, for which the results are partially in contrast with the standard models with uncorrelated measurements.

We refer the reader to the monographs of Rao [89], Lehmann and Casella [61] and Casella and Berger [11], which are pivotal sources for Section 2.1, and the monographs of Pázman [77] and Pukelsheim [87], which are essential for Sections 2.2 and 2.4.

2.1 Fisher information

Fisher information, or in the full wording "Fisher information that a random observable contains about the unknown parameter", named after Sir Ronald A. Fisher, is one of the most important concepts connected with parameter estimation and information theory, and plays a crucial role also in the presented Thesis.

There are several degrees of generality, in which we can handle the Fisher information. The needs of this thesis are limited to continuous random vectors depending on an unknown vector parameter, so we narrow the discussion to this class of models with the following definition:

Definition 14. *Let $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ be a probability density function of a continuous random vector $\boldsymbol{X}$ on $\mathbb{R}^n$ with respect to Lebesgue measure. Then the Fisher information matrix for unknown parameter $\boldsymbol{\theta}$ evaluated at prior guess $\boldsymbol{\theta}^*$ at the true value of the unknown parameter $\boldsymbol{\theta}$ is defined by*

$$\begin{aligned} \mathscr{I}(\boldsymbol{\theta}^*) &= \mathcal{E}_{\boldsymbol{X}} \left[\frac{\partial f(\boldsymbol{X} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial f(\boldsymbol{X} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \\ &= \int_{\boldsymbol{x} \in \mathbb{R}^n} \frac{\partial f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}} f(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\boldsymbol{x} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} . \end{aligned}$$

Besides Definition 14, which is common for the parameter estimation, Fisher information is also very popular in the information theory with strong motivation in physics. More

precisely, what physicist usually take as the Fisher information is its "entropy-like analogue" given by

$$\mathscr{I}(\boldsymbol{X}) = \mathcal{E}_{\boldsymbol{X}}\left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathsf{T}}}\right]; \tag{2.1}$$

see the monograph of Frieden [31] for more details. While in Definition 14 we measure how sensitive are the observations to the parameter, the physical formulation in (2.1) measures a distortion of the distribution.

Computing the expectation in Definition 14 might be challenging. Under some regularity conditions we can get a more convenient formula for the Fisher information.

Proposition 11. *Let $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ be a probability density function satisfying the following regularity conditions:*

i) $\boldsymbol{\Theta}$ *is open set,*

ii) the support of $f(\boldsymbol{x} \mid \boldsymbol{\theta})$, i.e. $\operatorname{supp}(f(\cdot \mid \boldsymbol{\theta})) = \{\boldsymbol{x} \mid f(\boldsymbol{x} \mid \boldsymbol{\theta}) > 0\}$, does not depend on $\boldsymbol{\theta}$,

iii) for any $\boldsymbol{x} \in \operatorname{supp}(f(\cdot \mid \boldsymbol{\theta}))$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the derivative $\frac{\partial f(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ exist,

iv) $\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbb{R}^n} f(\boldsymbol{x} \mid \boldsymbol{\theta})\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} \frac{\partial f(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathrm{d}\boldsymbol{x} = 0.$

Then

$$\mathscr{I}(\boldsymbol{\theta}^*) = -\mathcal{E}_{\boldsymbol{X}}\left[\frac{\partial^2 f(\boldsymbol{X} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathsf{T}}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

In particular, for the Gaussian random vector Mardia and Marshall [67] showed the following.

Proposition 12. *Let $\boldsymbol{X}$ be a Gaussian random vector depending on an unknown vector $\boldsymbol{\theta}$, with the vector of expected values $\mathcal{E}_{\boldsymbol{\theta}}[\boldsymbol{X}]$ and the variance-covariance matrix $\mathcal{V}_{\boldsymbol{\theta}}[\boldsymbol{X}]$. Then the Fisher information matrix is*

$$\mathscr{I}(\boldsymbol{\theta}^*) = \left(\frac{\partial \mathcal{E}_{\boldsymbol{\theta}}^{\mathsf{T}}[\boldsymbol{X}]}{\partial \boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}^{-1}[\boldsymbol{X}]\frac{\partial \mathcal{E}_{\boldsymbol{\theta}}[\boldsymbol{X}]}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} + \frac{1}{2}\operatorname{tr}\left\{\mathcal{V}_{\boldsymbol{\theta}}^{-1}[\boldsymbol{X}]\frac{\partial \mathcal{V}_{\boldsymbol{\theta}}[\boldsymbol{X}]}{\partial \boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}^{-1}[\boldsymbol{X}]\frac{\partial \mathcal{V}_{\boldsymbol{\theta}}[\boldsymbol{X}]}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right\}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

In statistical inference we can often summarize the information by taking a few key figures. That is, instead of a long list of observations, say $\boldsymbol{X}$, we take a statistic $\boldsymbol{T}(\boldsymbol{X})$, for which $\dim(\boldsymbol{T}(\boldsymbol{X})) \ll \dim(\boldsymbol{X})$. We may consider different statistics, but only some of them have the property that we do not lose any information, which is usually desired.

Definition 15. *A statistic $\boldsymbol{T}(\boldsymbol{X})$ is sufficient statistic for parameter $\boldsymbol{\theta}$ if the conditional distribution of the sample $\boldsymbol{X}$ given the value of $\boldsymbol{T}(\boldsymbol{X})$ does not depend on $\boldsymbol{\theta}$.*

Note that sufficient statistics depend solely on the observations and do not depend on unknown parameters. We can easily verify the sufficiency of a statistic by using the Fisher-Neymann Factorization Theorem:

Proposition 13 (Fisher-Neymann Factorization Theorem)**.** *Let $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ be a probability density function for random vector $\boldsymbol{X}$. A statistic $\boldsymbol{T}(\boldsymbol{X})$ is sufficient if and only if there exist functions $g(\boldsymbol{t} \mid \boldsymbol{\theta})$ and $h(\boldsymbol{x})$ such that $f(\boldsymbol{x} \mid \boldsymbol{\theta}) = g(\boldsymbol{T}(\boldsymbol{x}) \mid \boldsymbol{\theta})h(\boldsymbol{x})$.*

A simple example of a sufficient statistic is a bijective mapping, for which we have $f(\boldsymbol{x} \mid \boldsymbol{\theta}) = f(\boldsymbol{T}^{\mathrm{inv}}(\boldsymbol{T}(\boldsymbol{x})) \mid \boldsymbol{\theta}) = g(\boldsymbol{T}(\boldsymbol{x}) \mid \boldsymbol{\theta})$, where $\boldsymbol{T}^{\mathrm{inv}}$ is the inverse mapping of $\boldsymbol{T}$. The fact that sufficient statistics preserve the information has consequences concerning Fisher information.

Proposition 14. *Let $\boldsymbol{X}$ be a random vector, $\boldsymbol{T}(\boldsymbol{X})$ be a statistic, and let $\mathscr{I}_{\boldsymbol{X}}(\boldsymbol{\theta})$ and $\mathscr{I}_{\boldsymbol{T}(\boldsymbol{X})}(\boldsymbol{\theta})$ be the corresponding Fisher information matrices. Then $\mathscr{I}_{\boldsymbol{T}(\boldsymbol{X})}(\boldsymbol{\theta}) \preceq_{\mathrm{L}} \mathscr{I}_{\boldsymbol{X}}(\boldsymbol{\theta})$ with equality if and only if $\boldsymbol{T}(\boldsymbol{X})$ is a sufficient statistic.*

Fisher information is strongly connected with the quality of estimators in different ways. The first result we refer to is the Cramér-Rao bound:

Proposition 15 (Cramér-Rao bound)**.** *Let $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ be a probability density function of a random vector and let $\boldsymbol{T}(\boldsymbol{X})$ be any estimator of $\boldsymbol{\theta}$ satisfying*

$$\frac{\partial \mathcal{E}[\boldsymbol{T}(\boldsymbol{X})]}{\partial \boldsymbol{\theta}^{\mathsf{T}}} = \int_{\mathbb{R}^n} \frac{\partial}{\partial \boldsymbol{\theta}^{\mathsf{T}}} [\boldsymbol{T}(\boldsymbol{x}) f(\boldsymbol{x} \mid \boldsymbol{\theta})] \mathrm{d}\boldsymbol{x}, \text{ and } \mathcal{V}[\boldsymbol{T}(\boldsymbol{X})] < \infty.$$

Then

$$\mathcal{V}[\boldsymbol{T}(\boldsymbol{X})] \succeq_{\mathrm{L}} \frac{\partial \mathcal{E}^{\mathsf{T}}[\boldsymbol{T}(\boldsymbol{X})]}{\partial \boldsymbol{\theta}} \mathscr{I}_{\boldsymbol{X}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathcal{E}[\boldsymbol{T}(\boldsymbol{X})]}{\partial \boldsymbol{\theta}^{\mathsf{T}}}.$$

The origin of the Cramér-Rao bound is accredited to Cramér [14] and Rao [88], after who the result is named. It should be, however, mentioned [83, p510] that Fisher in his 1935 paper [29] already stated the Fisher information as the lower bound for the asymptotic variance. References for further extensions of the Cramér-Rao bound can be found in [61, Sec2.5–2.6].

The Cramér-Rao bound is only a lower bound on the quality of estimator (although upper bound would be more practical), and the variance of an estimator is exactly equal to the (matrix) inverse of Fisher information matrix if the estimator is efficient. Although in a general situation maximum likelihood estimator is not efficient for a finite sample, under some circumstances the inverse of the Fisher information is a good proxy for the estimator's variance. For example, we provide the following modified result of Pázman [79] (for the sake of simplicity we do not give a rigorous formulation):

Proposition 16 (Pázman [79])**.** *Let $\boldsymbol{X}$ be a Gaussian random vector with $\mathcal{E}[\boldsymbol{X}]$ depending on parameter vector $\boldsymbol{\theta}$. If the diagonal components of the variance matrix $\mathcal{V}[\boldsymbol{X}]$ are small, then the variance of the maximum likelihood estimator is close to the inverse of the Fisher information matrix evaluated at the (approximately) true value of the parameter, that is*

$$\mathcal{V}[\hat{\boldsymbol{\theta}}] \approx \mathscr{I}_{\boldsymbol{X}}^{-1}(\bar{\boldsymbol{\theta}}).$$

2.2 Experimental designs, information functions and efficiency

Under controlled experiments the measurements in individual trials of experiment depend on selected experimental conditions, which we formally call experimental design. More precisely:

Definition 16. *An exact n-point design on a set $\mathcal{X}$ of possible experimental conditions is a vector $\xi_n = (x_1, \ldots, x_n)$ with $x_i \in \mathcal{X}$, $i = 1, \ldots, n$. A set of all feasible exact n-point designs in the experiment is denoted by Ξ_n.*

A special type of experiments are those, where the experimental conditions are sampling times. In this case we use the term "sampling design" (see Definition 5), and we set $\xi_n = \boldsymbol{\tau}_n$, $\mathcal{X} = \mathcal{D}$ and $\Xi_n = \overline{\mathfrak{T}}_n$.

Under controlled experiments, the Fisher information matrix, which is a legitimate measure of quality of experiment, is a function of experimental conditions. That is, being given

Criterion	$\Phi[\mathscr{I}]$	Geometric description
D	$[\det(\mathscr{I})]^{1/m}$	Volume of the confidence ellipsoid Shannon information [13]
A	$[\frac{1}{m}\mathrm{tr}\{\mathscr{I}^{-1}\}]^{-1}$, $\mathscr{I} \in \mathbb{S}^m_{++}$ 0 otherwise	Average variance of estimates
E	$\lambda_{\min}(\mathscr{I})$	Maximal radius of the confidence ellipsoid
$\boldsymbol{c}$	$\boldsymbol{c}^\mathsf{T}\mathscr{I}^{-}\boldsymbol{c}$	Variance of $\boldsymbol{c}^\mathsf{T}\hat{\boldsymbol{\theta}}$

Table 2.1. Classical optimality criteria

a prior guess $\boldsymbol{\theta}^*$ at the true value of the unknown parameter $\boldsymbol{\theta}$ and an exact design ξ_n, we get the Fisher information matrix of the form $\mathscr{I}(\xi_n, \boldsymbol{\theta}^*)$.

A question to answer is a selection between two designs. An experimenter naturally prefers designs with "large" Fisher information matrix. While discrimination between two designs in a single-parameter model is straightforward, under a multiparameter setup the situation is different. A usual way for comparing two matrices is using the Loewner ordering. That is, a non-negative definite matrix $\mathscr{I}_1$ "Loewner dominates" non-negative definite matrix $\mathscr{I}_2$ if the matrix $\mathscr{I}_1 - \mathscr{I}_2$ is non-negative definite, which we denote by $\mathscr{I}_1 \succeq_\mathrm{L} \mathscr{I}_2$.

Since Loewner ordering on the set of all symmetric non-negative definite matrices $\mathbb{S}^n_+$ forms a partially ordered set, not all matrices are comparable. For that reason we define information functions:

Definition 17 ([87])**.** *An information function is any function $\Phi : \mathbb{S}^m_+ \mapsto [0, \infty)$ which is non-constant, concave, upper semicontinuous, positive homogeneous and Loewner isotonic (i.e., if $\mathscr{I}_1 \succeq_\mathrm{L} \mathscr{I}_2$ then $\Phi[\mathscr{I}_1] \geq \Phi[\mathscr{I}_2]$)*

An information function represents experimenter's preferences for the properties of the estimator – a criterion of experiment. In Table 2.1 we give the most popular classical optimality criteria, which conform to certain geometric or statistical properties. For more details we refer the reader to the monograph of Pázman [77] and Pukelsheim [87].

A generalization of the D-,E- and A-optimality criteria mentioned above is the class of Kiefer's Φ_p-optimality criteria [52, 51, 85, 87]. For a positive-definite matrix $\mathscr{I}$ we define the information function for Φ_p-optimality as

$$\Phi_p(\mathscr{I}) \equiv \begin{cases} (\frac{1}{m}\mathrm{tr}\{\mathscr{I}^p\})^{1/p} & \text{for } p \in (-\infty, 0) \cup (0, 1], \\ \det{}^{1/m}(\mathscr{I}) & \text{for } p = 0, \\ \lambda_{\min}(\mathscr{I}) & \text{for } p = -\infty, \end{cases}$$

and for a singular non-negative definite matrix $\mathscr{I}$ we define

$$\Phi_p(\mathscr{I}) \equiv \begin{cases} (\frac{1}{m}\mathrm{tr}\{\mathscr{I}^p\})^{1/p} & \text{for } p \in (0, 1], \\ 0 & \text{for } p \in [-\infty, 0]. \end{cases}$$

Definition 18. *We say that the design ξ^*_n is locally Φ-optimal under a prior guess $\boldsymbol{\theta}^*$ if*

$$\xi^*_n \in \arg\max_{\xi \in \Xi_n} \Phi[\mathscr{I}(\xi, \boldsymbol{\theta}^*)].$$

Obviously, for different values of $\boldsymbol{\theta}^*$ we obtain different optimal designs. Several extensions dealing with models, where the informativeness of experiment depends on the true

value of parameter can be found in the literature; see, e.g., the monograph of Pronzato and Pázman [84]

In applications we sometimes do not have to perform an experiment under optimal experimental conditions. For instance, we may prefer designs, which are close to an optimal design but possess other desirable properties. A common way how to measure the quality of a design is its efficiency.

Definition 19 ([87]). *The efficiency of a design ξ_n with respect to optimality criterion Φ under a prior guess $\boldsymbol{\theta}^*$ is*

$$\mathrm{eff}(\xi_n \mid \Phi, \boldsymbol{\theta}^*) \equiv \frac{\Phi[\mathscr{I}(\xi_n, \boldsymbol{\theta}^*)]}{\arg\max_{\zeta \in \Xi_n} \Phi[\mathscr{I}(\zeta, \boldsymbol{\theta}^*)]}.$$

We should note that although Φ denotes information functions, in many papers Φ represents any Loewner isotonic or antiisotonic functionals, which are to be maximised or minimised, respectively. For D-optimality, we can, e.g., maximise $\Phi[\mathscr{I}] = \det^{1/m}(\mathscr{I})$, but also minimise $\Phi[\mathscr{I}] = \ln\det(\mathscr{I})$. For the sake of simplicity, by Φ we will understand both situations and, when necessary, we declare this fact explicitly.

2.3 Regression experiments

Definition 20. *Let $\boldsymbol{\theta} \in \mathbb{R}^m$ be the unknown vector parameter, $\mathcal{X}$ be the set of feasible experimental conditions, $\eta_{\boldsymbol{\theta}}(x) : x \in \mathcal{X} \mapsto \mathbb{R}$ be the mean value of the observed random variable $y(x)$ and $\Sigma_{\boldsymbol{\theta}}(x_1, x_2)$ be a symmetric positive-definite function generating the variance-covariance matrix of the vector of observations $\boldsymbol{y}(\xi_n)$ (i.e., $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\xi_n^\mathsf{T}, \xi_n) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\xi_n) \in \mathbb{S}^n_{++}$ for all $\xi_n \in \Xi_n$), where η and Σ might depend on $\boldsymbol{\theta}$ or its subvector (displayed by a relevant subscript). Then by a regression experiment with the unknown parameter $\boldsymbol{\theta}$, mean value η, covariance function Σ and experimental designs belonging to Ξ_n, denoted by a quadruplet $(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}}, \Xi_n)$, we understand the model*

$$\begin{aligned} \boldsymbol{y}(\xi_n) &= \boldsymbol{\eta}_{\boldsymbol{\theta}}(\xi_n) + \boldsymbol{\varepsilon}(\xi_n), \\ \mathcal{E}[\boldsymbol{\varepsilon}(\xi_n)] &= \mathbf{0}_n \\ \mathcal{V}[\boldsymbol{\varepsilon}(\xi_n)] &= \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\xi_n). \end{aligned}$$

By a classical regression model we understand the situation $(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}}, \Sigma, \Xi_n)$, i.e., when $\frac{\partial \Sigma}{\partial \boldsymbol{\theta}^\mathsf{T}} = \mathbf{0}_m$. By $(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}}, \sigma^2_{\boldsymbol{\theta}}, \Xi_n)$ we denote the experiment $(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}}, \Xi_n)$ with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\xi_n) = \mathrm{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{\theta}}(\xi_n))$, where $\sigma^2_{\boldsymbol{\theta}}(x)$ is a positive function. If $\eta_{\boldsymbol{\theta}}(x) = \boldsymbol{f}^\mathsf{T}(x)\boldsymbol{\theta}$, where $\boldsymbol{f}(x)$ is a vector of regression functions, then the regression experiment is called linear, otherwise it is non-linear.

Note that the mean value and the covariance structure might have, in general, common parameters.

Let us consider the classical regression model $(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}}, \Sigma, \Xi_n)$. A usual way how to estimate $\boldsymbol{\theta}$ is using the least squares estimator:

$$\hat{\boldsymbol{\theta}}_{\mathbf{V}(\xi_n)} = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^m} \|\boldsymbol{y}(\xi_n) - \boldsymbol{\eta}_{\boldsymbol{\theta}}(\xi_n)\|^2_{\mathbf{V}(\xi_n)}, \tag{2.2}$$

where $\mathbf{V}(\xi_n)$ is a positive definite weight matrix depending on the design ξ_n. Additionally, if the distribution of the error vector $\boldsymbol{\varepsilon}(\xi_n)$ is Gaussian and $\mathbf{V}(\xi_n) = \boldsymbol{\Sigma}(\xi_n^\mathsf{T}, \xi_n) = \boldsymbol{\Sigma}(\xi_n)$, then the least squares estimator is equivalent to the maximum likelihood estimator.

For the classical linear regression experiments $(\boldsymbol{\theta}, \boldsymbol{f}^\mathsf{T}(x)\boldsymbol{\theta}, \Sigma, \Xi_n)$, we have

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{V}(\xi)} = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^m} \|\boldsymbol{y} - \boldsymbol{f}^\mathsf{T}(\xi_n)\boldsymbol{\theta}\|^2_{\mathbf{V}(\xi)} = [\boldsymbol{f}(\xi_n^\mathsf{T})\mathbf{V}^{-1}(\xi_n)\boldsymbol{f}^\mathsf{T}(\xi_n)]^- \boldsymbol{f}(\xi_n^\mathsf{T})\mathbf{V}^{-1}(\xi_n)\boldsymbol{y}(\xi_n). \tag{2.3}$$

To clarify the notation we recall that $\boldsymbol{f}(\xi_n^{\mathsf{T}}) = (\boldsymbol{f}(x_1), \ldots, \boldsymbol{f}(x_n)) = (\boldsymbol{f}^{\mathsf{T}}(\xi_n))^{\mathsf{T}}$.

It can be easily shown that for any $\boldsymbol{c} \in \mathscr{M}(\boldsymbol{f}(\xi_n^{\mathsf{T}}))$, we have that $\mathcal{E}[\boldsymbol{c}^{\mathsf{T}}\hat{\boldsymbol{\theta}}_{\mathbf{V}(\xi_n)}] = \boldsymbol{c}^{\mathsf{T}}\boldsymbol{\theta}$ and

$$\mathcal{V}[\boldsymbol{c}^{\mathsf{T}}\hat{\boldsymbol{\theta}}_{\mathbf{V}(\xi_n)}] = \boldsymbol{c}^{\mathsf{T}}[\boldsymbol{f}(\xi_n^{\mathsf{T}})\mathbf{V}^{-1}(\xi_n)\boldsymbol{f}^{\mathsf{T}}(\xi_n)]^{-}\boldsymbol{f}(\xi_n^{\mathsf{T}})\mathbf{V}^{-1}(\xi_n)\times \\ \boldsymbol{\Sigma}(\xi_n)\mathbf{V}^{-1}(\xi_n)\boldsymbol{f}^{\mathsf{T}}(\xi_n)[\boldsymbol{f}(\xi_n^{\mathsf{T}})\mathbf{V}^{-1}(\xi_n)\boldsymbol{f}^{\mathsf{T}}(\xi_n)]^{-}\boldsymbol{c}. \quad (2.4)$$

A fundamental result in the field of regression models is the Gauss-Markov theorem [77, 78, 87]:

Proposition 17 (Gauss-Markov theorem)**.** *Let $\boldsymbol{y}(\xi_n)$ satisfy classical linear regression model $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(x)\boldsymbol{\theta}, \Sigma, \Xi_n)$ (without a specified distribution of errors) and let $\hat{\boldsymbol{\theta}}_{\boldsymbol{\Sigma}(\xi_n)}$ be the least squares estimator defined by* (2.3)*. Then for any linear unbiased estimator $\boldsymbol{l}^{\mathsf{T}}\boldsymbol{y}$ of a linear function $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{\theta}$ we have*

$$\mathcal{V}[\boldsymbol{l}^{\mathsf{T}}\boldsymbol{y}] \geq \boldsymbol{c}^{\mathsf{T}}[\boldsymbol{f}(\xi_n^{\mathsf{T}})\boldsymbol{\Sigma}^{-1}(\xi_n)\boldsymbol{f}^{\mathsf{T}}(\xi_n)]^{-}\boldsymbol{c} = \mathcal{V}[\boldsymbol{c}^{\mathsf{T}}\hat{\boldsymbol{\theta}}_{\boldsymbol{\Sigma}(\xi_n)}],$$

with equality if and only if $\boldsymbol{l} = \boldsymbol{c}^{\mathsf{T}}[\boldsymbol{f}(\xi_n^{\mathsf{T}})\boldsymbol{\Sigma}^{-1}(\xi_n)\boldsymbol{f}^{\mathsf{T}}(\xi_n)]^{-}\boldsymbol{f}(\xi_n)\boldsymbol{\Sigma}^{-1}(\xi_n)$.

If the errors of the regression experiment are Gaussian, then $\boldsymbol{f}(\xi_n^{\mathsf{T}})\boldsymbol{\Sigma}^{-1}(\xi_n)\boldsymbol{f}^{\mathsf{T}}(\xi_n)$ is the Fisher information matrix. Henceforth, $\hat{\boldsymbol{\theta}}_{\boldsymbol{\Sigma}(\xi_n)}$ is an efficient estimator, cf. Cramér-Rao bound.

2.4 Basics of the theory of experimental design for regression models under uncorrelated observations

Consider an experiment with uncorrelated observations. From the properties of the Fisher information we get

$$\mathscr{I}(\xi_n, \boldsymbol{\theta}^*) = \sum_{i=1}^{n} \mathscr{I}(x_i, \boldsymbol{\theta}^*) = \sum_{x\in\mathcal{X}} n(x)\mathscr{I}(x, \boldsymbol{\theta}^*) = n\sum_{x\in\mathcal{X}} \frac{n(x)}{n}\mathscr{I}(x, \boldsymbol{\theta}^*), \quad (2.5)$$

where $n(x)$ is the number of trials under experimental condition x, that is, the Fisher information from experiment is equal to the sum of Fisher information matrices from individual trials.

Specifically, in the experiment $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(x)\boldsymbol{\theta}, \sigma^2, \Xi_n)$ we have

$$\mathscr{I}(x, \boldsymbol{\theta}^*) = \frac{\boldsymbol{f}(x)\boldsymbol{f}^{\mathsf{T}}(x)}{\sigma^2(x)}. \quad (2.6)$$

Analogously, for the non-linear experiment $(\boldsymbol{\theta}, \eta_{\boldsymbol{\theta}}, \sigma^2, \Xi_n)$, we have

$$\mathscr{I}(x, \boldsymbol{\theta}^*) = \frac{1}{\sigma^2(x)}\frac{\partial\eta_{\boldsymbol{\theta}}(x)}{\partial\boldsymbol{\theta}}\frac{\partial\eta_{\boldsymbol{\theta}}(x)}{\partial\boldsymbol{\theta}^{\mathsf{T}}}\bigg|_{\boldsymbol{\theta}^*}, \quad (2.7)$$

where $\boldsymbol{\theta}^*$ is a prior guess at the true value of the unknown parameter.

Definition 21. *A discrete measure $\xi(x) = n(x)/n$ is called an experimental design associated with exact design ξ_n.*

In the theory of optimal design of experiments associated experimental designs lead to a continuous relaxation; see the Kiefer's paper [50] and references therein.

Definition 22. *By approximate (or asymptotic) experimental design we understand an arbitrary discrete probability measure $\xi(x)$ with finite support on $\mathcal{X}$. By Ξ we denote the set of all approximate designs in experiment.*

The requirement for the finite support on $\mathcal{X}$ follows from the next proposition, which can be in different levels of generalization found in [12, 22, 28, 85]:

Proposition 18. *For any design measure $\xi \in \Xi$ exists a design measure $\zeta \in \Xi$ such that the cardinality of $\{x \in \mathcal{X} \mid \zeta(x) > 0\}$ is at most $\frac{1}{2}m(m+1)+1$, where m is the number of unknown parameters, and $\mathscr{I}(\xi, \boldsymbol{\theta}^*) = \mathscr{I}(\zeta, \boldsymbol{\theta}^*)$.*

While exact designs imply approximate design measures, the opposite is not true. If ξ is an approximate design, $n\xi(x)$ is not necessarily an integer. There, however, exist rounding methods, for which efficiency has been estimated [86].

Because of the positive homogenity of information functions, in the experiments with uncorrelated observations under the Fisher information matrix for approximate design ξ we understand

$$\mathscr{I}(\xi, \boldsymbol{\theta}^*) = \sum_{x \in \mathcal{X}} \mathscr{I}(x, \boldsymbol{\theta}^*)\xi(x). \tag{2.8}$$

Clearly, the set Ξ of all approximate designs and the set $\mathcal{M}_\Xi$ of all information matrices in experiment are convex. Moreover, if $\boldsymbol{f}(x)$ and $\sigma^2(x)$ are continuous and $\mathcal{X}$ is compact then $\mathcal{M}_\Xi$ is compact, too. We recall that an information function Φ is assumed to be concave and upper semicontinuous (that is $-\Phi$ is convex); therefore, the computation of optimal designs is a convex programming problem, which has important implications for existence of optimal design and computational aspects. For particular model assumptions this property was exploited by many authors, who used linear, semi-definite and other types of mathematical programming; see [105, 38, 95].

One of the main results of the presented theory is the Equivalence Theorem, which states necessary and sufficient conditions for optimality of approximate designs. The key is the concept of a directional derivative. The literature offers diverse definitions of the directional derivative. In the theory of experimental design, for convex set, given points Ω and Γ, the directional derivative of a (matrix/scalar) function $\boldsymbol{h}$ at the point Ω and in the direction Γ is

$$\partial \boldsymbol{h}(\Omega, \Gamma - \Omega) = \lim_{\Delta \searrow 0} \frac{\boldsymbol{h}[\Omega + \Delta(\Gamma - \Omega)] - \lim_{\Delta \searrow 0} \boldsymbol{h}[\Omega + \Delta(\Gamma - \Omega)]}{\Delta}. \tag{2.9}$$

Note that in the definition of the directional derivative as given above, we consider the directional continuity of the function $\boldsymbol{h}$ at the point Ω, which is important if $\boldsymbol{h}$ is not continuous at Ω and its behaviour in an arbitrary neighbourhood of Ω is analysed.

It follows from the convexity of Ξ and $\xi \mapsto -\Phi[\mathscr{I}(\xi, \boldsymbol{\theta}^*)]$ under experiments with uncorrelated observations that

Proposition 19. *If $\mathscr{I}(\xi, \boldsymbol{\theta}^*)$ is defined by* (2.8)*, then the design ξ^* is Φ-optimal if and only if for all $\xi \in \Xi$,*

$$\partial\Phi[\mathscr{I}(\xi^*, \boldsymbol{\theta}^*), \mathscr{I}(\xi, \boldsymbol{\theta}^*) - \mathscr{I}(\xi^*, \boldsymbol{\theta}^*)] \leq 0.$$

Moreover, if gradient of Φ, $\{\boldsymbol{\nabla}\Phi[\mathscr{I}]\}_{ij} = \frac{\partial\Phi[\mathscr{I}]}{\mathscr{I}_{ij}}$, exists then

$$\partial\Phi[\mathscr{I}(\xi^*, \boldsymbol{\theta}^*), \mathscr{I}(\xi, \boldsymbol{\theta}^*) - \mathscr{I}(\xi^*, \boldsymbol{\theta}^*)] = -\mathrm{tr}\{\boldsymbol{\nabla}\Phi[\mathscr{I}(\xi^*, \boldsymbol{\theta}^*)][\mathscr{I}(\xi^*, \boldsymbol{\theta}^*) - \mathscr{I}(\xi, \boldsymbol{\theta}^*)]\}.$$

Proposition 20 (Equivalence Theorem [52])**.** *Let $\mathscr{I}(\xi, \boldsymbol{\theta}^*)$ be defined by* (2.6) *and* (2.8)*, Φ be an information function and let $\boldsymbol{\nabla}\Phi$ exist. Then a design ξ^* is Φ-optimal if and only if*

$$\min_{x \in \mathcal{X}} \boldsymbol{f}^{\mathsf{T}}(x)\boldsymbol{\nabla}\Phi[\mathscr{I}(\xi^*, \boldsymbol{\theta}^*)]\boldsymbol{f}(x) = \mathrm{tr}\{\Phi[\mathscr{I}(\xi^*, \boldsymbol{\theta}^*)]\mathscr{I}(\xi^*, \boldsymbol{\theta}^*)\}.$$

Part II

Literature review

Chapter 3

Overview of the experimental design with correlated observations

The main aim of this chapter is to summarize the most important results in the field of experimental design for regression models subject to correlated observations. Dependent observations are typical for experiments related to stochastic processes; therefore, the assumed controllable variable is the time $t \in \mathcal{D}$, i.e., we consider sampling designs.

If not stated else, the model underlying this chapter is the classical linear regression model of the form

$$(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(t)\boldsymbol{\theta}, \Sigma, \overline{\mathfrak{T}}_n). \tag{3.1}$$

We recall that $\boldsymbol{\theta} \in \mathbb{R}^m$ is the unknown parameter with estimator given in (2.3), $\boldsymbol{f} : \mathcal{D} \mapsto \mathbb{R}^m$ is the vector of linearly independent regression functions, $\Sigma : \mathcal{D} \times \mathcal{D} \mapsto \mathbb{R}$ is the non-negative definite covariance function (or covariance kernel), and $\overline{\mathfrak{T}}_n$ is the set of designs with non-decreasing sampling times from $\mathcal{D} = [T_*, T^*]$, $T^* > T_* > 0$. By the Fisher information matrix in the model (3.1) we understand

$$\mathscr{I}(\boldsymbol{\tau}_n, \boldsymbol{\theta}^*) = \mathscr{I}(\boldsymbol{\tau}_n) = \boldsymbol{f}(\boldsymbol{\tau}_n^{\mathsf{T}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{f}^{\mathsf{T}}(\boldsymbol{\tau}_n). \tag{3.2}$$

In Section 3.1 we briefly present several important differences between regression models with and without correlated observations. From the experimental design point of view, the correlation yields computationally challenging optimisation problems. Unlike in the case of experimental design with uncorrelated errors, for construction of optimal designs for dependent observations we do not have a complete theory, and various authors either tried to build heuristic or approximate theories, or studied very specific models.

3.1 Aspects of experimental design for regression models with correlated observations

The presence of correlation between the measurements makes the relation between the selected design and the corresponding amount of information less transparent.

In the theory of approximate experimental design with uncorrelated observations, for a given support of design points, the resulting Fisher information matrix is a linear function of weights (for both approximate and exact designs it takes an additive form) and the mapping $\xi \mapsto -\Phi[\mathscr{I}(\xi)]$ is convex. In contrast with this approximate theory presented in Section 2.4, the Fisher information matrix (3.2) is not additive and is less transparent, the

mapping $\boldsymbol{\tau}_n \mapsto -\Phi[\mathscr{I}(\boldsymbol{\tau}_n, \boldsymbol{\theta}^*)]$ is not convex on $\overline{\mathfrak{T}}_n$, and the set of all information matrices in the experiment $\mathcal{M}_{\overline{\mathfrak{T}}_n} = \{\mathscr{I}(\boldsymbol{\tau}_n, \boldsymbol{\theta}^*) \mid \boldsymbol{\tau}_n \in \overline{\mathfrak{T}}_n\}$ is not convex as well.

Additionally, in many instances the set of competing designs is, because of the model properties, reduced to the non-compact set $\mathfrak{T}_n \subset \overline{\mathfrak{T}}_n$, or $\mathfrak{T}_n$ is assumed directly. As a result, optimal designs need not exist, and such cases are not exceptional.

The dependence between observations leads to less apparent acquiring of the information. An illustrative example was given by Pázman [80]: if the observations in the regression model (3.1) are uncorrelated, i.e., $\Sigma(t_i, t_j) = 0$ if and only if $t_i \neq t_j$, we can easily identify those design points, which give zero information. Following the formula (2.6) we obtain that such points satisfy equation $\boldsymbol{f}(x) = \mathbf{0}_m$. Now, consider a single-parameter model $(\theta, f(t)\theta, \Sigma, \mathfrak{T}_n)$ with $f(t_1) = 0$, $f(t_2) = 1$, $\Sigma(t_1, t_1) = \Sigma(t_2, t_2) = 1$ and $\Sigma(t_1, t_2) \neq 0$. If we perform one observation at t_2 only, then the amount of information is equal to 1. On the other hand, if we make the measurements at t_1 and t_2, then the Fisher information is equal to $[1 - \Sigma^2(t_1, t_2)]^{-1} > 1$. In particular, if $\Sigma(t_1, t_2) \in \{-1, 1\}$ then the information is infinite, because we can compute the value of unknown parameter exactly.

Proposition 21 (Pázman [80] motivated by Näther [71])**.** *Let us consider the regression model $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(t)\boldsymbol{\theta}, \Sigma, \mathfrak{T}_n)$. Let $\boldsymbol{\tau}_n \in \mathfrak{T}_n$ be a design, $\boldsymbol{a}(\nu) = \sum_{t \in \boldsymbol{\tau}_n} [\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)]^{-1}_{\nu,t} \boldsymbol{f}(t)$ for $\nu \in \boldsymbol{\tau}_n$, and $\boldsymbol{\theta}$ be fixed. If $\boldsymbol{a}(t_0) = \mathbf{0}_m$ for some t_0, then t_0 provides zero contribution to the information about $\boldsymbol{\theta}$.*

Proposition 22 (Pázman [80])**.** *Let us consider a model $(\boldsymbol{\beta}, \boldsymbol{f}^{\mathsf{T}}(t)\boldsymbol{\theta}, \Sigma_{\boldsymbol{\beta}}, \mathfrak{T}_n)$ with Gaussian errors. Let $\boldsymbol{\tau}_n \in \mathfrak{T}_n$ be a design, $\boldsymbol{b}(\nu_1, \nu_2) = \sum_{t,s \in \boldsymbol{\tau}_n} [\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)]^{-1}_{\nu_1,t} \frac{\partial \boldsymbol{\Sigma}(t,s)}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)]^{-1}_{s,\nu_2}$ for $\nu_1, \nu_2 \in \boldsymbol{\tau}_n$, and $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ be fixed. If there is a subvector $\tilde{\boldsymbol{\tau}} \subseteq \boldsymbol{\tau}_n$ such that $\boldsymbol{b}(t_{01}, t_{02}) = \mathbf{0}_m$ for all $t_{01} \in \boldsymbol{\tau}_n \setminus \tilde{\boldsymbol{\tau}}$ and $t_{02} \in \boldsymbol{\tau}_n$, then the design points $\boldsymbol{\tau}_n \setminus \tilde{\boldsymbol{\tau}}$ provide zero contribution to the information about $\boldsymbol{\beta}$.*

Statistical relationship between measurements also has implications on the attainable amount of information. Due to formula (2.5), if there exists a design point with nonzero associated design measure and positive definite individual Fisher information matrix, then for the increasing number of measurements, the amount of information increases above any limits. On the other hand, this is not always true for models with correlated observations. If $\{\boldsymbol{\tau}^{(n)}\}_n \in \mathfrak{C}_{\mathcal{D}}$ is a sequence of designs covering $\mathcal{D}$, then the Fisher information matrix $\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\theta}^*)$ converges to a matrix $\mathscr{I}_\infty(\boldsymbol{\tau}^*) = \lim_{n\to\infty} \mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\theta}^*)$ with the largest eigenvalue bounded. Consequently, the maximum likelihood estimator is not consistently estimable; see, e.g., the papers of Crowder [15], Sweeting [103], and the references therein.

Boundedness of the Fisher information matrix (3.2) (in the Loewner ordering sense) in the presence of correlation also has some practical aspects. Pázman [79] proposed a modification of the efficiency ratio in Definition 19, where the denominator is substituted for the value $\Phi[\tilde{\mathscr{I}}]$ with $\tilde{\mathscr{I}}$ being any suitable reference matrix. In particular, under correlated observations, Pázman [79] further noted that a suitable choice for the reference matrix $\tilde{\mathscr{I}}$ is the asymptotic Fisher information matrix $\mathscr{I}_\infty(\boldsymbol{\theta}^*)$ obtained by observation of the trajectory of the stochastic process at every time from the experimental domain $\mathcal{D}$. To distinguish between the conventional efficiency in Definition 19 and the efficiency defined by Pázman, for the latter we use the nomenclature "ultimate efficiency" suggested by Harman [37].

Definition 23 (Ultimate efficiency)**.** *The ultimate efficiency of a design $\boldsymbol{\tau}$ with respect to optimality criterion Φ under a prior guess $\boldsymbol{\theta}^*$ is the ratio*

$$\operatorname{ueff}(\boldsymbol{\tau} \mid \Phi, \boldsymbol{\theta}^*) \equiv \frac{\Phi[\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\theta}^*)]}{\Phi[\mathscr{I}_\infty(\boldsymbol{\theta}^*)]}.$$

The evidence shows that for a small number of observations, the ultimate efficiency might be high also for non-optimal designs. If this is the case, then the need for a further

optimisation of the experiment is questionable. Additionally, Sack and Ylvisaker [92] highlighted the use of ultimate efficiency in deciding how many observations the experimenter should take and whether the costs are appropriate to the information gain if we performed another observation.

3.2 Asymptotic approach using reproducing kernel Hilbert spaces

Probably the first systematic analysis of the optimal experimental designs for the models of the form (3.1), on which we focus in this section, was published by Sacks and Ylvisaker [92, 93] in 1960's. Before this period, we can only find results for specific models subject to dependent observations like those by Hoel [41, 42], who studied optimal designs for general variance of the least squares estimator for polynomial regression.

Since obtaining of explicit finite sample solutions to experimental design problems for (3.1) is achievable only exceptionally, to circumvent this difficulty, Sacks and Ylvisaker built their results on asymptotic arguments using the theory of the reproducing kernel Hilbert spaces (RKHSs) – a popular and elegant approach to analysis of time series and stochastic processes; see the papers of Parzen [75, 74]. From 1960's to 1980's, the RKHSs and the approach of Sacks and Ylvisaker have been the motivation and leading tool also in other publications concerning various models and problems, of which we can point, for instance, Wahba [106, 107], Sacks and Ylvisaker [94] or Eubank, Smith and Smith [23, 24, 25].

3.2.1 Single-parameter setup

Firstly consider a single-parameter variant of model (3.1) with experimental designs from $\mathfrak{T}_n$, i.e., we assume the observations to satisfy the model $(\theta, f(t)\theta, \Sigma, \mathfrak{T}_n)$. We take the sampling interval $\mathcal{D} = [0, T^*]$

According to the Gauss-Markov Theorem, for any n-point design $\boldsymbol{\tau}_n \in \mathfrak{T}_n$, the variance of the efficient linear estimator is equal to the scalar value $[\boldsymbol{f}^\mathsf{T}(\boldsymbol{\tau}_n)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{f}(\boldsymbol{\tau}_n)]^{-1} = \mathscr{I}^{-1}(\boldsymbol{\tau}_n) = \|\boldsymbol{f}(\boldsymbol{\tau}_n)\|^{-2}_{\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)} = \|f\|_{\boldsymbol{\tau}_n}$. As usual, by an exactly optimal sampling design we understand such vector $\boldsymbol{\tau}_n^* \in \mathfrak{T}_n$ such that

$$[\boldsymbol{f}^\mathsf{T}(\boldsymbol{\tau}_n^*)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n^*)\boldsymbol{f}(\boldsymbol{\tau}_n^*)]^{-1} \leq [\boldsymbol{f}^\mathsf{T}(\boldsymbol{\tau}_n)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{f}(\boldsymbol{\tau}_n)]^{-1} \text{ for all } \boldsymbol{\tau}_n \in \mathfrak{T}_n.$$

Alternatively, for a single-parameter setup, Sacks and Ylvisaker [92] proposed the following asymptotic "solution":

Definition 24. *(Asymptotic optimality for single-parameter setup) We say that a sequence of designs $\{\boldsymbol{\tau}^{*(n)}\}_n$ covering $\mathcal{D}$ is asymptotically optimal, if*

$$\lim_{n\to\infty} \frac{\mathscr{I}^{-1}(\boldsymbol{\tau}^{*(n)}) - \mathscr{I}_\infty^{-1}}{\min_{\boldsymbol{\tau}\in\mathfrak{T}_n} \mathscr{I}^{-1}(\boldsymbol{\tau}) - \mathscr{I}_\infty^{-1}} = 1$$

Evidently, such definition puts some restrictions on the choice of the regression function f. Let us define $\|f\|_{\mathfrak{F}_n} \equiv \sup_{\boldsymbol{\tau}\in\mathfrak{T}_n} \|f\|_{\boldsymbol{\tau}}$ and $\|f\|_{\mathfrak{F}} \equiv \lim_{n\to\infty} \sup_{\boldsymbol{\tau}_n\in\mathfrak{T}_n} \|f\|_{\boldsymbol{\tau}_n}$. In the present setup, for each n, we assume that the regression function f belongs to the set

$$\mathfrak{F}_n \equiv \{f : \mathcal{D} \mapsto \mathbb{R} \mid \|f\|_{\mathfrak{F}_n} < \infty\}.$$

Since $\sup_{\boldsymbol{\tau}_n\in\mathfrak{T}_n} \|f\|_{\boldsymbol{\tau}_n}$ is non-decreasing in n, $\mathfrak{F}_n \supset \mathfrak{F}_{n+1}$, and we define

$$\mathfrak{F} \equiv \{f : \mathcal{D} \mapsto \mathbb{R} \mid \|f\|_{\mathfrak{F}} < \infty\}.$$

Similarly to $\|f\|_{\mathfrak{F}}$, we can define the inner product $\langle\cdot,\cdot\rangle_{\mathfrak{F}}$.

Proposition 23 ([92]). *$\mathfrak{F}_n$ with the norm $\|\cdot\|_{\mathfrak{F}_n}$ forms a Banach space. $\mathfrak{F}$ with the norm $\|\cdot\|_{\mathfrak{F}}$ forms a reproducing kernel Hilbert space associated with Σ, i. e., for each $t \in \mathcal{D}$, $\Sigma(\cdot, t) \in \mathfrak{F}$, and for each $f \in \mathfrak{F}$ and each $t \in \mathcal{D}$, we can write $\langle f, \Sigma(\cdot, t)\rangle_{\mathfrak{F}} = f(t)$.*

Proposition 24 ([92]). *Let $\hat{\theta}_n$ be any sequence of the least squares estimates for θ, where n is the number of observations. Then there exists a random variable $\hat{\theta}_\infty$ such that $\mathcal{E}[\hat{\theta}_n - \hat{\theta}_\infty] \to 0$ and $\hat{\theta}_\infty$ is an unbiased estimator satisfying $\mathcal{V}[\hat{\theta}_\infty] = \|f\|_{\mathfrak{F}}^{-2}$, where $\|f\|_{\mathfrak{F}}^{-2}$ is taken as zero if $f \notin \mathfrak{F}$.*

Let for the underlying stochastic process, say $\{X(t)\}_{t\geq 0}$, hold true the equality $X(t_2) \mid X(t_1) = X(t_1)$ almost surely if $t_2 = t_1$. Then, obviously, replicated sampling times give no information about the parameters of the process, and for any boundary design $\tilde{\boldsymbol{\tau}}_n$, i.e., $\tilde{\boldsymbol{\tau}}_n \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$, there exists a design $\boldsymbol{\tau}_n \in \mathfrak{T}_n$ such that $\mathscr{I}(\boldsymbol{\tau}_n) \succeq_{\mathrm{L}} \mathscr{I}(\tilde{\boldsymbol{\tau}}_n)$ (we give a detailed discussion in Part III). In some instances, we can find sequences of n-point sampling designs $\{\boldsymbol{\tau}_n^{(k)}\}_k$ with $\lim_{k\to\infty} \boldsymbol{\tau}_n^{(k)} = \tilde{\boldsymbol{\tau}}_n$, where the amount of information grows as $k \to \infty$ and $\lim_{k\to\infty} \mathscr{I}(\boldsymbol{\tau}_n^{(k)}) \neq \mathscr{I}(\tilde{\boldsymbol{\tau}}_n)$, and so there is no optimal design. Such difficulties arise, when the Fisher information matrix is not continuous at the boundary points belonging to $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$. To ensure the existence of optimal sampling designs, it is sufficient to find conditions yielding the continuity of $\mathscr{I}(\boldsymbol{\tau})$ on $\overline{\mathfrak{T}}_n$.

For the model $(\theta, f(t)\theta, \Sigma, \mathfrak{T}_n)$, Sacks and Ylvisaker [92] built their arguments on the fact that $\mathfrak{F}$ with the norm $\|\cdot\|_{\mathfrak{F}}$ forms a reproducing kernel Hilbert space. The idea is to rewrite the Fisher information $\|f\|_{\boldsymbol{\tau}_n}^2$ as $\|\mathbf{P}_{\boldsymbol{\tau}_n} f\|_{\mathfrak{F}}^2$, where $\mathbf{P}_{\boldsymbol{\tau}_n}$ is the projection operator defined on $\mathfrak{F}$ to the subspace of functions $\mathscr{M}(\{\Sigma(\cdot, t) \mid t \in \boldsymbol{\tau}_n\})$. The continuity of the Fisher information $\|f\|_{\boldsymbol{\tau}_n}^2$ in $\boldsymbol{\tau}_n$ on $\overline{\mathfrak{T}}_n$ is, hence, equivalent to the continuity of the mapping $\boldsymbol{\tau}_n \mapsto \mathbf{P}_{\boldsymbol{\tau}_n} f$ on $\overline{\mathfrak{T}}_n$.

We can alternatively define the continuity in terms of convergence of functional spaces as follows: Consider a a functional space $\mathfrak{G}_\epsilon(\boldsymbol{\tau}_n) = \mathscr{M}(\{\Sigma(\cdot, t) \mid t \in (t_i - \epsilon, t_i + \epsilon), i = 1, \ldots, n\})$. Then for any fixed $\boldsymbol{\tau}_n$, the space $\mathfrak{G}_\epsilon(\boldsymbol{\tau}_n)$ is continuous at $\epsilon = 0$ if $\cap_{\epsilon>0}\mathfrak{G}_\epsilon(\boldsymbol{\tau}_n) = \mathscr{M}(\{\Sigma(\cdot, t) \mid t \in \boldsymbol{\tau}_n\})$. Because of the isomorphism between $\mathscr{M}(\{\Sigma(\cdot, t) \mid t \in \boldsymbol{\tau}_n\})$ and Hilbert space of vectors with the norm $\|\cdot\|_{\boldsymbol{\tau}_n}$, the continuity of $\mathfrak{G}_\epsilon(\boldsymbol{\tau}_n)$ at $\epsilon = 0$ for all $\boldsymbol{\tau}_n \in \overline{\mathfrak{T}}_n$ implies the continuity of $\|f\|_{\boldsymbol{\tau}_n}^2$ in $\boldsymbol{\tau}_n$ on $\overline{\mathfrak{T}}_n$.

Definition 25 ([92]). *The covariance function Σ is said to have a simple present if in $\mathfrak{F}$,*

$$\bigcap_{\epsilon>0} \mathscr{M}(\{\Sigma(\cdot, t) \mid t \in (t_i - \epsilon, t_i + \epsilon), i = 1, \ldots, n\}) = \mathscr{M}(\{\Sigma(\cdot, t) \mid t \in \boldsymbol{\tau}_n\})$$

for all $\boldsymbol{\tau}_n = (t_1, \ldots, t_n)^{\mathsf{T}} \in \overline{\mathfrak{T}}_n$ and all n.

Proposition 25 ([92]). *If Σ has a simple present and if $\{\boldsymbol{\tau}_n^{(k)}\}_{k=1}^\infty$ is any sequence from $\mathfrak{T}_n$ with limit $\tilde{\boldsymbol{\tau}}_n \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$, then $\mathbf{P}_{\boldsymbol{\tau}_n^{(k)}} f \to \mathbf{P}_{\tilde{\boldsymbol{\tau}}_n} f$ for all $f \in \mathfrak{F}$.*

Proposition 26 ([92]). *Consider the model $(\theta, f(t)\theta, \Sigma, \mathfrak{T}_n)$, where*

$$f(t) = \int_{\mathcal{D}} \Sigma(s, t)\varphi(s)\mathrm{d}s$$

and φ is a continuous function on $\mathcal{D}$. Let the covariance function Σ satisfy the following assumptions:

i) $\Sigma(s, t)$ is continuous on $\mathcal{D} \times \mathcal{D}$ and has continuous derivatives up to order two at every (s, t) with $s \neq t$. At (s, t) with $s = t$ we assume that Σ has all right and left derivatives up to order two.

ii) For $t \in (0, T^*)$, $\alpha(t) = \lim_{s \nearrow t} \frac{\partial \Sigma(s,t)}{\partial t} - \lim_{s \searrow t} \frac{\partial \Sigma(s,t)}{\partial t}$ *is continuous on* $(0, T^*)$, *and* $\inf_{t \in (0,T^*)} \alpha(t) > 0$ *and* $\sup_{t \in (0,T^*)} \alpha(t) < \infty$, *so that* $\alpha(t)$ *can be extended to* $\mathcal{D}$.

iii) For each $t_0 \in \mathcal{D}$ *the function* $\lim_{t \searrow t_0} \frac{\partial^2 \Sigma(\cdot,t)}{\partial t^2} \in \mathfrak{F}$, *and* $\sup_{t_0 \in \mathcal{D}} \left\| \lim_{t \searrow t_0} \frac{\partial^2 \Sigma(\cdot,t)}{\partial t^2} \right\|_{\mathfrak{F}} < \infty$.

Then

$$\lim_{n \to \infty} n^2 \inf_{\boldsymbol{\tau}_n \in \mathfrak{T}_n} \|f - \mathbf{P}_{\boldsymbol{\tau}_n} f\|_{\mathfrak{F}}^2 = \frac{1}{12} \left(\int_{\mathcal{D}} [\alpha(t) \varphi^2(t)]^{1/3} \mathrm{d}t \right)^3.$$

Furthermore, the sequence of designs $\{\boldsymbol{\tau}^{*(n)}\}_n$ *with the* i*th component* $t_i^{*(n)}$ *of* $\boldsymbol{\tau}^{*(n)}$ *defined by the smallest solution to the equation*

$$\int_0^{t_i^{*(n)}} [\alpha(t) \varphi^2(t)]^{1/3} \mathrm{d}t = \frac{i-1}{n-1} \int_{\mathcal{D}} [\alpha(t) \varphi^2(t)]^{1/3} \mathrm{d}t, \ i = 1, \ldots, n,$$

is an asymptotically optimal sequence of designs.

3.2.2 Multi-parameter setup

Given the model (3.1) with a vector parameter $\boldsymbol{\theta}$, the situation is similar to the case discussed in the previous subsection. The assumptions i) – iii) of Proposition 26 remain valid for multivariate parameter, and the regression functions $f_i(t)$'s, which belong to the reproducing kernel Hilbert space formed by $\mathfrak{F}$ with $\| \cdot \|_{\mathfrak{F}}$ induced by the inner product $\langle \cdot, \cdot \rangle_{\mathfrak{F}}$, satisfy

$$f_i(t) = \int_{\mathcal{D}} R(s,t) \varphi_i(s) \mathrm{d}s, \ t \in \mathcal{D}, \ i = 1, \ldots, m. \tag{3.3}$$

Proposition 27 ([93]). *Let the assumptions i) – iii) of Proposition 26 be satisfied, and* $\boldsymbol{\varphi}(t) = (\varphi_1(t), \ldots, \varphi_m(t))^\mathsf{T}$ *be linearly independent. Then* $\boldsymbol{f}(t) = (f_1(t), \ldots, f_m(t))^\mathsf{T}$ *are linearly independent over* $\mathcal{D}$.

The ijth element of the asymptotic Fisher information matrix $\mathscr{I}_\infty$, $i, j = 1, \ldots, m$, is $\{\mathscr{I}_\infty\}_{ij} = \langle f_i, f_j \rangle_{\mathfrak{F}}$. Using the terminology of the previous subsection, we can write the Fisher information matrix for any n-point design $\boldsymbol{\tau}_n \in \overline{\mathfrak{T}}_n$ as $\{\mathscr{I}(\boldsymbol{\tau}_n)\}_{ij} = \langle \mathbf{P}_{\boldsymbol{\tau}_n} f_i, \mathbf{P}_{\boldsymbol{\tau}_n} f_j \rangle$. Due to this fact, we can derive a similar result for existence of optimal sampling to those in Proposition 25 [93].

Proposition 28 ([93]). *Under the conditions of Proposition 27, the asymptotic Fisher information matrix is positive definite.*

In contrast with the single-parameter setup of the previous subsection, in this section we compare matrices, for which we employ the information functions [87], which we discussed in Section 2.2. Following the Definition 24 we define the asymptotic optimality for mutli-parameter setup:

Definition 26 ([93]). *Let* Φ *be a given information function. We say that a sequence of designs* $\{\boldsymbol{\tau}^{*(n)}\}_n$ *covering* $\mathcal{D}$ *is asymptotically*

- $\Phi 1$*-optimal, if*

$$\lim_{n \to \infty} \frac{\inf_{\boldsymbol{\tau}_n \in \mathfrak{T}_n} \Phi[\mathscr{I}^{-1}(\boldsymbol{\tau}_n)] - \Phi[\mathscr{I}_\infty^{-1}]}{\Phi[\mathscr{I}^{-1}(\boldsymbol{\tau}^{*(n)})] - \Phi[\mathscr{I}_\infty^{-1}]} = 1,$$

- $\Phi 2$-*optimal, if*
$$\lim_{n\to\infty} \frac{\Phi[\mathscr{I}_\infty] - \sup_{\boldsymbol{\tau}_n\in\mathfrak{T}_n} \Phi[\mathscr{I}(\boldsymbol{\tau}_n)]}{\Phi[\mathscr{I}_\infty] - \Phi[\mathscr{I}(\boldsymbol{\tau}^{*(n)})]} = 1,$$
- $\Phi 3$-*optimal, if*
$$\lim_{n\to\infty} \inf_{\boldsymbol{\tau}_n\in\mathfrak{T}_n} \frac{\Phi[\mathscr{I}^{-1}(\boldsymbol{\tau}_n) - \mathscr{I}_\infty^{-1}]}{\Phi[\mathscr{I}^{-1}(\boldsymbol{\tau}^{*(n)}) - \mathscr{I}_\infty^{-1}]} = 1,$$
- $\Phi 4$-*optimal, if*
$$\lim_{n\to\infty} \inf_{\boldsymbol{\tau}_n\in\mathfrak{T}_n} \frac{\Phi[\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}_n)]}{\Phi[\mathscr{I}_\infty - \mathscr{I}^{-1}(\boldsymbol{\tau}^{*(n)})]} = 1.$$

A special position amongst the information functions considered in the paper by Sacks and Ylvisaker for the purpose of the asymptotic optimality has $\Phi[\mathscr{I}] = \mathrm{tr}\{\mathbf{M}\mathscr{I}\}$, where $\mathbf{M}$ is a fixed non-negative definite matrix. The reason is two-fold. Firstly, for $\Phi[\mathbf{M}-\mathbf{A}]$ and $\mathbf{A}$ close to the zero matrix $\mathbf{0}_{m\times m}$, we have $\Phi[\mathbf{M}-\mathbf{A}] = \Phi[\mathbf{M}] - \mathrm{tr}\{\boldsymbol{\nabla}\Phi[\mathbf{M}]\mathbf{A}\} + \mathrm{o}(\|\mathbf{A}\|)$, where $\|\mathbf{A}\|$ is a matrix norm of $\mathbf{A}$ and $\{\boldsymbol{\nabla}\Phi[\mathbf{M}]\}_{ij} = \frac{\partial\Phi[\mathscr{I}]}{\partial\mathscr{I}_{ij}}(\mathbf{M})$. Thus, $\Phi[\mathscr{I}(\boldsymbol{\tau}_n)] = \Phi[\mathscr{I}_\infty - (\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}_n))] = \Phi[\mathscr{I}_\infty] - \mathrm{tr}\{\boldsymbol{\nabla}\Phi[\mathscr{I}_\infty](\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}_n))\} + \mathrm{o}(\|\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}_n)\|)$, which yields

$$\Phi[\mathscr{I}_\infty] - \Phi[\mathscr{I}(\boldsymbol{\tau}_n)] = \mathrm{tr}\{\boldsymbol{\nabla}\Phi[\mathscr{I}_\infty]\mathscr{I}_\infty\} - \mathrm{tr}\{\boldsymbol{\nabla}\Phi[\mathscr{I}_\infty]\mathscr{I}(\boldsymbol{\tau}_n)\} + \mathrm{o}(\|\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}_n)\|). \quad (3.4)$$

An analogous formula can be obtained for $\Phi[\mathscr{I}^{-1}(\boldsymbol{\tau}_n)]$.

Secondly, for some of the selected criteria of asymptotic optimality under information functions defined by $\mathrm{tr}\{\mathbf{M}\mathscr{I}\}$, $\mathrm{tr}\{\mathbf{M}\mathscr{I}^{-1}\}$ or their supremums through $\mathbf{M}$ on appropriate sets are equivalent to asymptotic optimality of some standard criteria.

The idea for analysis of asymptotically optimal designs is based on the evidence of Theorem 3.1 by Sacks and Ylvisaker [92] (cf. Proposition 26) that one can express designs as uniformly distributed quantiles of a certain density function, say h. That is, for a sequence of designs $\{\boldsymbol{\tau}^{(n)}\}_n$, $\boldsymbol{\tau}^{(n)} = (t_0^{(n)}, t_1^{(n)}, \ldots, t_n^{(n)}) \in \mathfrak{T}_{n+1}$, the ith component is the smallest solution to the equation

$$\int_0^{t_i^{(n)}} h(t)\mathrm{d}t = i/n,$$

with the convention that $t_0^{(n)} = 0$ and $t_n^{(n)} = T^*$. Such sequences are called regular sequences generated by h, denoted by $\mathfrak{RS}(h)$.

Proposition 29 ([93]). *Let $\{\boldsymbol{\tau}^{(n)}\}_n$ be a $\mathfrak{RS}(h)$, $\alpha(t)$ be defined in the assumption ii) of Proposition 26 and φ is related to f through equation* (3.3)*. Then for*

$$\lim_{n\to\infty} n^2\|f - \mathbf{P}_{\boldsymbol{\tau}^{(n)}} f\|_{\mathfrak{F}}^2 = \frac{1}{12}\int_{\mathcal{D}} \alpha(t)\varphi^2(t),$$

any of the following conditions is sufficient:

i) $\int_{\mathcal{D}}(h^2(t))^{-1}\mathrm{d}t < \infty$ *and* φ *is a continuous function,*

ii) φ/h *is a continuous function,*

iii) there exists a constant K *such that for all* $a, b \in \mathcal{D}$, $(b-a)\int_a^b h^2(t)\mathrm{d}t \le K\left(\int_a^b h(t)\mathrm{d}t\right)^2$.

Proposition 30 ([93]). *Let $f_1, \dots, f_m$ be regression functions of the form* (3.3) *with associated $\varphi_1, \dots, \varphi_m$, and let $a_1, \dots, a_m$ be positive numbers. If $\{\boldsymbol{\tau}^{(n)}\}_n$ is any sequence of designs covering $\mathcal{D}$, then*

$$\liminf_{n\to\infty} n^2 \sum_{k=1}^{m} \|f_k - \mathbf{P}_{\boldsymbol{\tau}^{(n)}} f_k\|_{\mathfrak{F}}^2 \geq \frac{1}{12} \left[\int_{\mathcal{D}} \left(\alpha(t) \sum_{k=1}^{m} a_k \varphi_k^2(t) \right)^{1/3} \mathrm{d}t \right]^3,$$

with equality if $\{\boldsymbol{\tau}^{(n)}\}_n$ is a $\mathfrak{RS}(h)$, where

$$h(t) = \frac{\left[\alpha(t) \sum_{k=1}^{m} a_k \varphi_k^2(t)\right]^{1/3}}{\int_{\mathcal{D}} \left[\alpha(t) \sum_{k=1}^{m} a_k \varphi_k^2(t)\right]^{1/3} \mathrm{d}t} = \frac{\left[\alpha(t) \boldsymbol{\varphi}^{\mathsf{T}}(t) \mathrm{diag}(\boldsymbol{a}) \boldsymbol{\varphi}(t)\right]^{1/3}}{\int_{\mathcal{D}} \left[\alpha(t) \boldsymbol{\varphi}^{\mathsf{T}}(t) \mathrm{diag}(\boldsymbol{a}) \boldsymbol{\varphi}(t)\right]^{1/3} \mathrm{d}t}.$$

The previous proposition is the key for several results on asymptotic optimality. In this regard we denote

$$h^*_{\mathbf{C}}(t) \equiv \frac{\left[\alpha(t) \boldsymbol{\varphi}^{\mathsf{T}}(t) \mathbf{C} \boldsymbol{\varphi}(t)\right]^{1/3}}{\int_{\mathcal{D}} \left[\alpha(t) \boldsymbol{\varphi}^{\mathsf{T}}(t) \mathbf{C} \boldsymbol{\varphi}(t)\right]^{1/3} \mathrm{d}t},$$

where $\mathbf{C}$ is a non-negative definite matrix.

In the sequel we give several statements on asymptotic optimality. A more detailed discussion of these results is provided in the original paper of Sacks and Ylvisaker [93].

Proposition 31 ([93]). *Let $\Phi[\mathscr{I}] = \mathrm{tr}\{\mathbf{M}\mathscr{I}\}$, where $\mathbf{M}$ is a non-negative definite matrix. If $\{\boldsymbol{\tau}^{*(n)}\}_n$ is a $\mathfrak{RS}(h^*_{\mathbf{M}})$, then $\{\boldsymbol{\tau}^{*(n)}\}_n$ is asymptotically $\Phi 1$- and $\Phi 4$-optimal.*

Proposition 32 ([93]). *Suppose Φ is a strict and continuously differentiable criterion so that $\Phi[\mathscr{I}_\infty] - \Phi[\mathscr{I}(\boldsymbol{\tau}_n)] = \mathrm{tr}\{\mathbf{M}(\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}_n))\} + \mathrm{o}(\|\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}_n)\|)$, where $\mathbf{M}$ is a positive-definite matrix (cf. analogy with* (3.4)*). Then $\{\boldsymbol{\tau}^{*(n)}\}_n$, which is a $\mathfrak{RS}(h^*_{\mathbf{M}})$, is asymptotically $\Phi 2$-optimal. Specially, for $\Phi_{\mathrm{D}}[\mathscr{I}] = \det(\mathscr{I})$ (equivalent to D-optimality criterion, but not information function), if $\{\boldsymbol{\tau}^{*(n)}\}_n$ is a $\mathfrak{RS}(h^*_{\mathscr{I}_\infty^{-1}})$, then $\{\boldsymbol{\tau}^{*(n)}\}_n$ is asymptotically $\Phi_{\mathrm{D}} 2$-optimal.*

Proposition 33 ([93]). *Let $\Phi[\mathscr{I}] = \min_{\mathbf{M}\in\mathscr{K}} \mathrm{tr}\{\mathbf{M}\mathscr{I}\}$, where $\mathscr{K}$ is a compact set of non-negative matrices. Let $\mathbf{M}^*$ minimize $\int_{\mathcal{D}} \left[\alpha(t) \boldsymbol{\varphi}^{\mathsf{T}}(t) \mathbf{M} \boldsymbol{\varphi}(t)\right]^{1/3} \mathrm{d}t$ over $\mathscr{K}$. Then, if $\{\boldsymbol{\tau}^{*(n)}\}_n$ is a $\mathfrak{RS}(h^*_{\mathbf{M}^*})$, then $\{\boldsymbol{\tau}^{*(n)}\}_n$ is asymptotically $\Phi 4$-optimal.*

Proposition 34 ([93]). *Let $\Phi[\mathscr{I}] = \det(\mathscr{I})$. Let $\mathscr{K} = \{\mathbf{M} \mid \mathbf{M} \succ \mathbf{0}_{m\times m}, \det(\mathbf{M}) = 1\}$ and let $\mathbf{M}^*$ maximise $\int_{\mathcal{D}} \left[\alpha(t) \boldsymbol{\varphi}^{\mathsf{T}}(t) \mathbf{M} \boldsymbol{\varphi}(t)\right]^{1/3} \mathrm{d}t$ over $\mathscr{K}$ (even though $\mathscr{K}$ is not compact, it could be shown that the minimum is attained, and thus we can take the closure of $\mathscr{K}$). If $\{\boldsymbol{\tau}^{*(n)}\}_n$ is a $\mathfrak{RS}(h^*_{\mathbf{M}^*})$, then $\{\boldsymbol{\tau}^{*(n)}\}_n$ is asymptotically $\Phi 4$-optimal.*

Proposition 35 ([93]). *Let $\Phi[\mathscr{I}] = \mathrm{tr}\{\mathbf{M}\mathscr{I}\}$, where $\mathbf{M}$ is a non-negative definite matrix. Let $\{\boldsymbol{\tau}^{*(n)}\}_n$ be a $\mathfrak{RS}(h^*_{\mathscr{I}_\infty^{-1}\mathbf{M}\mathscr{I}_\infty^{-1}})$. If $\mathscr{I}(\{t \mid h^*_{\mathscr{I}_\infty^{-1}\mathbf{M}\mathscr{I}_\infty^{-1}}(t) > 0\})$ is non-singular, then $\{\boldsymbol{\tau}^{*(n)}\}_n$ is asymptotically $\Phi 1$- and $\Phi 3$-optimal.*

Proposition 36 ([93]). *If Φ is strict and continuously differentiable so that $\Phi[\mathscr{I}^{-1}(\boldsymbol{\tau}_n)] - \Phi[\mathscr{I}_\infty^{-1}]$ $= \mathrm{tr}\{\mathbf{M}(\mathscr{I}^{-1}(\boldsymbol{\tau}_n)] - \mathscr{I}_\infty^{-1})\} + \mathrm{o}(\|\mathscr{I}^{-1}(\boldsymbol{\tau}_n)] - \mathscr{I}_\infty^{-1}\|)$ for some positive definite matrix $\mathbf{M}$, then a regular sequence $\{\boldsymbol{\tau}^{*(n)}\}_n$ generated by $h_{\mathscr{I}_\infty^{-1}\mathbf{M}\mathscr{I}_\infty^{-1}}$ is asymptotically $\Phi 1$-optimal.*

Proposition 37 ([93]). *Let $\mathscr{K}$ be a compact set of non-negative definite matrices and $\Phi[\mathscr{I}] =$ $\mathrm{tr}\{\mathbf{M}\mathscr{I}\}$. Let $\mathbf{M}^*$ maximise $\int_{\mathcal{D}} \left[\alpha(t) \boldsymbol{\varphi}^{\mathsf{T}}(t) \mathscr{I}_\infty^{-1} \mathbf{M} \mathscr{I}_\infty^{-1} \boldsymbol{\varphi}(t)\right]^{1/3} \mathrm{d}t$ over $\mathscr{K}$. If $\{\boldsymbol{\tau}^{*(n)}\}_n$ is a $\mathfrak{RS}(h^*_{\mathscr{I}^*_\infty \mathbf{M}^* \mathscr{I}^*_\infty})$ and if $\lim_{n\to\infty} \mathscr{I}(\boldsymbol{\tau}^{*(n)})$ is non-singular, then $\{\boldsymbol{\tau}^{*(n)}\}_n$ is asymptotically $\Phi 3$-optimal.*

Proposition 38 ([93]). *If* $\mathscr{K} = \{\mathbf{M} \mid \mathbf{M} \succ \mathbf{0}_{m\times m}, \det(\mathbf{M}) = 1\}$ *and* $\Phi(\mathscr{I}) = \det(\mathscr{I})$, *then there is a matrix* $\mathbf{M}^* \in \mathscr{K}$ *which maximises* $\int_{\mathcal{D}} \left[\alpha(t)\boldsymbol{\varphi}^{\mathsf{T}}(t)\mathscr{I}_\infty^{-1}\mathbf{M}\mathscr{I}_\infty^{-1}\boldsymbol{\varphi}(t)\right]^{1/3} \mathrm{d}t$ *over* $\mathscr{K}$ *and* $h^*_{\mathscr{I}^*_\infty \mathbf{M}^* \mathscr{I}^*_\infty}$ *generates a regular sequence* $\{\boldsymbol{\tau}^{*(n)}\}_n$ *which is asymptotically* $\Phi 3$- *and* $\Phi 4$-*optimal.*

Proposition 39 ([93]). *Let* $\mathscr{K}$ *be a compact set of non-negative definite matrices and* $\overline{\mathscr{K}}$. *Let* $\Phi(\mathscr{I}) = \max_{\mathbf{M}\in\mathscr{K}}\{\mathbf{M}\mathscr{I}\} = \max_{\mathbf{M}\in\overline{\mathscr{K}}}\{\mathbf{M}\mathscr{I}\}$. *Let* $\mathbf{M}^*$ *maximise* $\int_{\mathcal{D}} \left[\alpha(t)\boldsymbol{\varphi}^{\mathsf{T}}(t)\mathbf{M}\boldsymbol{\varphi}(t)\right]^{1/3} \mathrm{d}t$ *and let* $\{\boldsymbol{\tau}^{*(n)}\}_n$ *be a* $\mathfrak{RS}(h^*_{\mathbf{M}^*})$. *If, for all* $\mathbf{M} \in \mathscr{K}$,

$$n^2 \mathrm{tr}\{\mathbf{M}(\mathscr{I}_\infty - \mathscr{I}(\boldsymbol{\tau}^{*(n)}))\} \le \frac{1}{12}\int_{\mathcal{D}} \alpha(t)\boldsymbol{\varphi}^{\mathsf{T}}(t)\mathbf{M}\boldsymbol{\varphi}(t)[h^*_{\mathbf{M}^*}(t)]^{-2}\mathrm{d}t + \mathrm{o}(1),$$

then $\{\boldsymbol{\tau}^{*(n)}\}_n$ *is asymptotically* $\Phi 4$-*optimal.*

3.3 Alternative asymptotic theories

A usual criticism towards explicit optimal designs for best linear unbiased estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{\Sigma}(\boldsymbol{\tau})}$, cf. (2.3), under the model (3.1) with the information matrix (3.2) is that the solutions might strongly depend on the covariance function, which might be misspecified. Bickel and Herzberg [6] suggested to estimate the unknown parameter by using ordinary least squares estimator $\hat{\boldsymbol{\theta}}_{\mathbf{I}_n}$, where, for an n-point design $\boldsymbol{\tau}_n \in \overline{\mathfrak{T}}_n$, the corresponding variance of the estimator is

$$\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}_n} \mid \boldsymbol{\tau}_n] = [\boldsymbol{f}(\boldsymbol{\tau}_n^{\mathsf{T}})\boldsymbol{f}^{\mathsf{T}}(\boldsymbol{\tau}_n)]^{-1}\boldsymbol{f}(\boldsymbol{\tau}_n^{\mathsf{T}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{f}^{\mathsf{T}}(\boldsymbol{\tau}_n)[\boldsymbol{f}(\boldsymbol{\tau}_n^{\mathsf{T}})\boldsymbol{f}^{\mathsf{T}}(\boldsymbol{\tau}_n)]^{-1} \tag{3.5}$$

cf. (2.4). Note that $\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}_n} \mid \boldsymbol{\tau}_n]$ is not the Fisher information matrix.

Although $\hat{\boldsymbol{\theta}}_{\mathbf{I}_n}$ is not a minimum variance estimator, the argument for use is a small difference in loss of efficiency [17, 111] and may possess another good properties [6, 70].

The asymptotics studied in papers of Bickel and Herzberg [6] and [7] is of a different nature than the one studied by Sacks and Ylvisaker [92, 93]. More precisely, while Sacks and Ylvisaker assumed a bounded experimental domain (normalised to the unit interval $[0, 1]$) with a fixed covariance function $\Sigma(\cdot, \cdot)$, Bickel and Herzberg considered a symmetric experimental domain of the form $[-nT^*, nT^*]$ and autocorrelation function $\gamma\rho[n(t_i - t_j)]$, where $\rho(\cdot)$ is a positive definite function such that $\rho(0) = 1$ and $\lim_{z\to\infty}\rho(z) = 0$. The approach of Bickel and Herzberg is not very practical in applications: for instance, in a cancer-research experiment the observation time domain is limited due to ethical reasons.

Nevertheless, the use of the variance of the estimator $\hat{\boldsymbol{\theta}}_{\mathbf{I}_n}$ given by (3.5) became a motivation for further research of asymptotic methods in experimental design with dependent observations provided by Dette, Pepelyshev and Zhigljavsky [17] and Zhigljavsky, Dette and Pepelyshev [111], who followed approximate design theory [52] (cf. Section 2.4).

The main idea of Dette, Pepelyshev and Zhigljavsky is straightforward: Let $\xi(t)$ be any probability measure on $\mathcal{D}$ and let us denote

$$\mathbf{M}(\xi) = \int_{\mathcal{D}} \boldsymbol{f}(t)\boldsymbol{f}^{\mathsf{T}}(t)\mathrm{d}\xi(t)$$

and

$$\mathbf{B}(\xi, \xi) = \int_{\mathcal{D}\times\mathcal{D}} \Sigma(t, s)\boldsymbol{f}(t)\boldsymbol{f}^{\mathsf{T}}(s)\mathrm{d}\xi(t)\mathrm{d}\xi(s).$$

Then for the variance-covariance matrix of the linear estimator $\hat{\boldsymbol{\theta}}_{\mathbf{I}}$ we have

$$\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi] \propto \mathbf{M}^{-1}(\xi)\mathbf{B}(\xi, \xi)\mathbf{M}^{-1}(\xi).$$

Here, $\xi \in \Xi$, where Ξ is the set of all approximate designs on $\mathcal{D}$, and we want to minimise the measure of variance $\Phi(\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi])$ with respect to ξ. (In the context of the notation in Section 2.2, if Φ is an information function, then we want to maximise $\Phi(\mathcal{V}^{-1}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi])$. For the reader it is, however, more practical if we keep the original notation of the paper [17]). Hence we say that an approximate design ξ^* is Φ-optimal, if, for all $\xi \in \Xi$,

$$\Phi(\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi^*]) \leq \Phi(\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi]).$$

We should note that the presented approach requires the designs ξ not to be supported by those subsets of $\mathcal{D}$ with $\boldsymbol{f}(t) = \mathbf{0}_m$, which is conceptually different to the suggestions of Pázman [80] (cf. Section 3.1) concerning the amount of information contained in the observations.

Although there is an attempt to incorporate convexity into the optimisation problem (the set Ξ), the mapping $\xi \mapsto \Phi(\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi])$ is not convex in general, so the main difficulty raised by the presence of correlation is still actual. On the other hand, the fact that we select designs from a convex set Ξ enabled Dette et al. [17] to comfortably use the concept of the directional derivative (see, e.g., Section 2.4) which yielded the following proposition:

Proposition 40 ([17]). *Let ξ^* be any approximate design minimising $\Phi(\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi])$. Then the inequality*

$$\phi(t, \xi^*) \leq b(t, \xi^*) \tag{3.6}$$

holds for all $t \in \mathcal{D}$, where

$$\begin{aligned} \phi(t, \xi) &= \boldsymbol{f}^{\mathsf{T}}(t)\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi]\boldsymbol{\nabla}\Phi(\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi])\mathbf{M}^{-1}(\xi)\boldsymbol{f}(t), \\ b(t, \xi) &= \operatorname{tr}\{\boldsymbol{\nabla}\Phi(\mathcal{V}[\hat{\boldsymbol{\theta}}_{\mathbf{I}} \mid \xi])\mathbf{M}^{-1}(\xi)\mathbf{B}(\xi, \xi_t)\mathbf{M}^{-1}(\xi)\}, \end{aligned}$$

and ξ_t is a probability measure concentrated at t. Moreover, (3.6) *is an equality for ξ^*-almost all t.*

A specific formulations of necessary conditions for D-optimality and $\boldsymbol{c}$-optimality can be found in [17].

Proposition 41 ([17]). *Consider the regression model $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(t), \boldsymbol{\theta}, \Sigma, \Xi)$, and let*

$$\boldsymbol{g}_{\xi}(t) = \int_{\mathcal{D}} \Sigma(t, s)\boldsymbol{f}(s)\mathrm{d}\xi(s) - \mathbf{B}(\xi, \xi)\mathbf{M}^{-1}(\xi)\boldsymbol{f}(t). \tag{3.7}$$

i) If $\boldsymbol{g}_{\xi^*}(t) = 0$ for all $t \in \mathcal{D}$, then the design ξ^* is universally optimal (i.e., with respect to all antiisotonic functionals Φ).

ii) If the design ξ^* is universally optimal, then the function $\boldsymbol{g}_{\xi^*}(t)$ can be represented in the form $\boldsymbol{g}_{\xi^*}(t) = \gamma(t)\boldsymbol{f}(t)$, where $\gamma(t)$ is a non-negative function such that $\gamma(t) = 0$ for all t belonging to the support of ξ.

Dette et al. [17] studied also some specific models, where the necessary and sufficient conditions did not require a detailed specifications of covariance functions. In general, this theory however requires some specifications of Σ, which is not completely in line with the original idea to construct optimal designs without the knowledge of covariance structure.

3.4 The virtual noise method of Pázman and Müller

The virtual noise method, which was described by Pázman and Müller [81] is one of the most prominent alternatives to computation of optimal design subject to correlated errors.

The idea of the method is based on a perturbation of observations by a supplementary (virtual) noise. Given the regression experiment $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(t)\boldsymbol{\theta}, \sigma^2 K, \mathfrak{T}_n)$, where K is a known positive definite function, we contaminate the observations by the white noise $\pi_{\xi,\gamma}(t)$ with the variance

$$\mathcal{V}[\pi(t)] = \sigma^2 \gamma \ln\left(\frac{\xi_{\max}}{\xi(t)}\right) = \begin{cases} +\infty & , \xi(t) = 0 \\ 0 & , \xi(t) = \xi_{\max} \\ \to 0 & , \gamma \to 0 \text{ and } \xi(t) > 0 \end{cases}.$$

Here, $\xi(t)$ is any approximate design on $\mathcal{D}$ and $\xi_{\max} = \max_{\mathcal{D}} \xi(t)$. The constant $\gamma > 0$ is a perturbation parameter, which needs to be small.

Now, let $\boldsymbol{\tau}_\xi = \mathrm{supp}\xi$, that is, $\boldsymbol{\tau}_\xi$ is a sampling design consisting of sampling times supporting the probability measure ξ. The Fisher information matrix of the perturbed problem is then

$$\mathscr{I}_{(\gamma)}(\xi) = \frac{1}{\sigma^2} \boldsymbol{f}(\boldsymbol{\tau}_\xi^{\mathsf{T}}) \left[\mathbf{K}(\boldsymbol{\tau}_\xi) + \gamma \mathrm{diag}\left\{\ln\left(\frac{\xi_{\max}}{\xi(t)}\right)\right\}\right]^{-1} \boldsymbol{f}^{\mathsf{T}}(\boldsymbol{\tau}_\xi). \tag{3.8}$$

The perturbed Fisher information matrix has the following properties:

i) Let $\{\xi_n(t)\}_{n\in\mathbb{N}}$ be any sequence of designs with common support and $\lim_{n\to\infty} \xi_n(t) = \xi(t)$. Then for each $\gamma > 0$, $\lim_{n\to\infty} \mathscr{I}_{(\gamma)}(\xi_n) = \mathscr{I}_{(\gamma)}(\xi)$.

ii) Let $\xi_{\boldsymbol{\tau}}$ be a uniform design with $\boldsymbol{\tau} = \mathrm{supp}\xi$. Then $\mathscr{I}_{(\gamma)}(\xi_{\boldsymbol{\tau}}) = \mathscr{I}(\boldsymbol{\tau}_\xi)$.

iii) For all ξ and $\gamma > 0$, $\mathscr{I}_{(\gamma)}(\xi) \preceq_{\mathrm{L}} \mathscr{I}(\boldsymbol{\tau}_\xi)$.

The rationale for the virtual noise method is then given by the following proposition

Proposition 42 ([81]). *Let $\Xi_{(n)}$ be the set of all approximate designs with support of no more than n points, n and $\gamma \geq 0$ be fixed, and suppose that $\min_{\boldsymbol{\tau}_p \in \mathfrak{T}_p : p \leq n} -\Phi[\mathscr{I}(\boldsymbol{\tau}_p)] < \infty$. Then*

i) if

$$\xi^* \in \arg\min_{\xi\in\Xi_{(n)}} -\Phi[\mathscr{I}_{(\gamma)}(\xi)], \tag{3.9}$$

then also $\boldsymbol{\tau} = \xi_{\mathrm{supp}\xi}$ (i.e., uniform design on $\mathrm{supp}\xi$) solves (3.9) *and*

$$\boldsymbol{\tau} \in \arg\min_{\boldsymbol{\tau}_p \in \mathfrak{T}_p : p \leq n} -\Phi[\mathscr{I}(\boldsymbol{\tau}_p)]. \tag{3.10}$$

ii) if $\boldsymbol{\tau}^*$ solves (3.10), *then $\xi_{\boldsymbol{\tau}^*}$ solves* (3.9).

Problem (3.9) still requires a minimization under the constraint $\boldsymbol{\tau}_p \in \mathfrak{T}_p : p \leq n$ which is discontinuous. We can avoid this constraint if we cut the design measure ξ at the level $1/n$. Then problem (3.10) corresponds to a minimization without constraints, cf [81].

Chapter 4

Experimental design for processes described by stochastic differential equations

In this section we review the results contributed to the design of experiments for processes described by stochastic differential equations.

We start this chapter with a discussion on the so-called product covariance structures. Even though this topic fits better the general overview in the previous chapter, it is listed in the present chapter due to the strong relation to some stochastic differential equations.

We recall that if any process $\{X(t)\}_{t\geq 0}$ is described by some stochastic differential equation, then we understand that $\{X(t)\}_{t\geq 0}$ is its weak solution.

4.1 Models with product covariance structures

Consider a regression model $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(t)\boldsymbol{\theta}, \Sigma, \mathfrak{T}_n)$, where

$$\Sigma(s,t) = \sigma^2 u(\min\{s,t\})v(\max\{s,t\}), \tag{4.1}$$

and where σ^2 is the only unknown factor of the covariance structure. Such instances are known as models with product covariance structures, which contain, for instance, Brownian motions, Brownian bridges, and stationary and nonstationary Ornstein-Uhlenbeck processes. Note that we assume designs without replications.

The usual obstacles with design problems with correlated observations is the fact that the Fisher information matrix depends on the inverse of the covariance matrix: $\mathscr{I}(\boldsymbol{\tau}_n) = \mathscr{I}(\boldsymbol{\tau}_n, \boldsymbol{\theta}^*) = \boldsymbol{f}(\boldsymbol{\tau}_n^{\mathsf{T}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{f}^{\mathsf{T}}(\boldsymbol{\tau}_n)$, cf. (3.2). We can, however, find several notes on the properties of the matrices with product structure and their inverses. For instance, Gantmacher [34] provided

$$\det(\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)) = u(t_1)v(t_n)\prod_{i=2}^{n}[u(t_i)v(t_{i-1}) - u(t_{i-1})v(t_i)],$$

hence $\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)$ is regular for $\boldsymbol{\tau}_n \in \mathfrak{T}_n$ if $u(t)/v(t)$ is strictly increasing, and Karlin [49] revealed the Jacobi structure of $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)$, i.e., the inverse is a tridiagonal matrix.

Besides some highlights by Sacks and Ylvisaker [92], the available literature offers two mainstreams in the use of product covariance structure in the theory of experimental design. The first are the results of Mukherjee [69] on estimation of the path (not the unknown

parameters) of a second order zero mean random process by minimising integrated mean square error or maximal mean square error.

The second contribution, which is more relevant for the presented thesis, is that of Harman and Štulajter [39, 40] and Harman [37]. The key result here is the explicit formula for the Fisher information matrix:

Proposition 43 ([39]). *Let the observations follow the model* $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(t)\boldsymbol{\theta}, \Sigma, \mathfrak{T}_n)$ *with* Σ *defined in* (4.1)*, and let* $\boldsymbol{\tau}_n$ *be a design. Then*

$$\sigma^2\{\mathscr{I}(\boldsymbol{\tau}_n)\}_{ij} = \frac{f_i(t_1)f_j(t_1)}{u(t_1)v(t_1)} + \sum_{k=2}^{n} \frac{\left(\frac{f_i(t_k)}{v(t_k)} - \frac{f_i(t_{k-1})}{v(t_{k-1})}\right)\left(\frac{f_j(t_k)}{v(t_k)} - \frac{f_j(t_{k-1})}{v(t_{k-1})}\right)}{\left(\frac{u(t_k)}{v(t_k)} - \frac{u(t_{k-1})}{v(t_{k-1})}\right)}.$$

Harman [37] showed the following relation between processes with product covariance structure and Brownian motions:

Proposition 44 ([37]). *Let* $r(t) = u(t)/v(t)$ *be strictly increasing such that* $r(t) \to 0$ *to* 0 *as* $t \to \infty$ *and* $r(t) \to +\infty$ *as* $t \to \infty$*, and denote* $\tilde{\boldsymbol{\tau}}_n^* = \boldsymbol{r}(\boldsymbol{\tau}_n^*)$*,* $\tilde{\mathcal{D}}^* = [r(T_*), r(T^*)]$ *and* $\tilde{\mathfrak{T}}_n$ *be the set of the corresponding sampling designs, and* $\tilde{\boldsymbol{f}}(t) = \boldsymbol{f}(r^{-1}(t))/v(t)$*. Then the design* $\boldsymbol{\tau}_n^*$ *is an optimal design for the regression model* $(\boldsymbol{\theta}, \boldsymbol{f}^{\mathsf{T}}(t)\boldsymbol{\theta}, \Sigma, \mathfrak{T}_n)$ *if and only if* $\tilde{\boldsymbol{\tau}}_n^*$ *is an optimal design for the regression model* $(\boldsymbol{\theta}, \tilde{\boldsymbol{f}}^{\mathsf{T}}(t)\boldsymbol{\theta}, \sigma^2 \min\{\cdot,\cdot\}, \tilde{\mathfrak{T}}_n)$.

The covariance structure $\min\{\cdot,\cdot\}$ corresponds to the standard Wiener process. Additionally, note that the continuity of the Fisher information matrix yields existence of optimal design belonging to $\mathfrak{T}_n$.

4.2 Nonstationary Ornstein-Uhlenbeck processes

Suppose we can observe a nonstationary Ornstein-Uhlenbeck process $\{X(t)\}_{t\geq 0}$ governed by the stochastic differential equation

$$\mathrm{d}X(t) = \theta_1[\theta_2 - X(t)]\mathrm{d}t + \sigma \mathrm{d}W(t),\ X(0) = X_0 \text{ fixed}, \tag{4.2}$$

where $\{W(t)\}_{t\geq 0}$ is the standard Wiener process.

It is a well-known fact that we can write

$$X(t) = \mathrm{e}^{-\theta_1 t}X_0 + (1 - \mathrm{e}^{-\theta_1 t})\theta_2 + \varepsilon(t),$$

where $\varepsilon(t)$ is a Gaussian random process with the product covariance structure ($t_2 \geq t_1$)

$$\begin{aligned} \mathcal{C}[X(t_1), X(t_2)] &= \frac{\sigma^2}{2\theta_1} u(t_1)v(t_2), \\ u(t) &= (\mathrm{e}^{\theta_1 t} - \mathrm{e}^{-\theta_1 t}), \\ v(t) &= \mathrm{e}^{-\theta_1 t}. \end{aligned}$$

Optimal sampling of experiments for the nonstationary Ornstein-Uhlenbeck process was to a certain extend studied by Harman and Štulajter [39], who assumed X_0 and θ_2 to be unknown parameters of interest, while θ_1 was a known constant.

For an n-point design $\boldsymbol{\tau}_n$, the corresponding regression model is

$$\boldsymbol{X}(\boldsymbol{\tau}_n) = \begin{pmatrix} \mathrm{e}^{-\theta_1 t_1} & 1 - \mathrm{e}^{-\theta_1 t_1} \\ \vdots & \vdots \\ \mathrm{e}^{-\theta_1 t_n} & 1 - \mathrm{e}^{-\theta_1 t_n} \end{pmatrix} \begin{pmatrix} X_0 \\ \theta_2 \end{pmatrix} + \boldsymbol{\varepsilon}(\boldsymbol{\tau}_n) = \left(\mathbf{e}^{\boldsymbol{\theta}_1 \boldsymbol{\tau}_n}, \mathbf{1}_n - \mathbf{e}^{\boldsymbol{\theta}_1 \boldsymbol{\tau}_n}\right) \begin{pmatrix} X_0 \\ \theta_2 \end{pmatrix} + \boldsymbol{\varepsilon}(\boldsymbol{\tau}_n).$$

The product covariance structure of the corresponding regression model yields the Fisher information matrix for $(X_0, \theta_2)^\mathsf{T}$

$$\mathscr{I}(\boldsymbol{\tau}_n) = \frac{2\theta_1}{\sigma^2} \begin{pmatrix} (1-\mathrm{e}^{-2\theta_1 t_1})^{-1} & (1+\mathrm{e}^{-\theta_1 t_1})^{-1} \\ (1+\mathrm{e}^{-\theta_1 t_1})^{-1} & \tanh\left(\frac{\theta_1}{2} t_1\right) + \sum_{i=2}^{n} \tanh\left(\frac{\theta_1}{2}(t_i - t_{i-1})\right) \end{pmatrix}$$

Since hyperbolic tangent $\tanh(z)$ a concave function on the positive halfline, $\sum_{i=2}^{n} \tanh(z_i)$ is a Schur concave function and thus its optimum is attained for some equivalued vector $z_i \equiv z^*$. Further,

$$\sum_{i=2}^{n} \tanh(z_i) = (n-1) \sum_{i=2}^{n} \frac{1}{n-1} \tanh(z_i) \leq (n-1) \tanh\left(\sum_{i=1}^{n} z_i/(n-1) \right).$$

Therefore, for a fixed t_1, $t_1 < T^*$, the Loewner optimal sampling is equidistant with the corresponding Fisher information matrix [39]

$$\mathscr{I}(t_1) = \frac{2\theta_1}{\sigma^2} \begin{pmatrix} (1-\mathrm{e}^{-2\theta_1 t_1})^{-1} & (1+\mathrm{e}^{-\theta_1 t_1})^{-1} \\ (1+\mathrm{e}^{-\theta_1 t_1})^{-1} & \tanh\left(\frac{\theta_1}{2} t_1\right) + (n-1) \tanh\left(\frac{\theta_1}{2} \cdot \frac{T^* - t_1}{n-1}\right) \end{pmatrix}.$$

Consequently, given the information function, the objective function $\Phi[\mathscr{I}(t_1)]$ is reduced to one dimension.

4.3 Stationary Ornstein-Uhlenbeck processes

For the time tending to infinity, distribution of the nonstationary Ornstein-Uhlenbeck process $\{X(t)\}_{t\geq 0}$ that solves stochastic differential equation (4.2) with $\theta_1 > 0$, converges to some stationary distribution.

We say that $\{Y(t)\}_{t\geq 0} = \lim_{\nu\to\infty}\{X(\nu+t)\}_{\nu+t\geq 0}$ is the stationary Ornstein-Uhlenbeck process. For $\{Y(t)\}_{t\geq 0}$ we have

$$\begin{aligned} \mathcal{E}[Y(t)] &= \lim_{\nu\to\infty} \mathcal{E}[X(\nu+t)] = \theta_2, \\ \mathcal{C}[Y(t), Y(t+s)] &= \lim_{\nu\to\infty} \mathcal{C}[X(\nu+t), X(\nu+t+s)] = \frac{\sigma^2}{2\theta_1} \mathrm{e}^{-\theta_1 s}. \end{aligned}$$

Note that $\mathcal{C}[Y(t), Y(t+s)] = \frac{\sigma^2}{2\theta_1} \mathrm{e}^{\theta_1 t} \mathrm{e}^{-\theta_1(t+s)} = u(t)v(t+s)$, where $u(t) = \frac{\sigma^2}{2\theta_1}\mathrm{e}^{\theta_1 t}$ and $v(t) = \mathrm{e}^{-\theta_1 t}$. That is, the covariance structure is of the product form.

Design of experiments for stationary Ornstein-Uhlenbeck processes were studied by Kiseľák and Stehlík [53] and Zagoraiou and Antognini [110], who, in contrast with the formulation implied by the nonstationary Ornstein-Uhlenbeck process (4.2), took the ratio $\frac{\sigma^2}{2\theta_1}$ as one parameter.

Proposition 45 ([53, 110])**.** *Let the mean value parameter θ_2 be the only unknown parameter of the stationary Ornstein-Uhlenbeck process. Then the equidistant design is an optimal design for estimation.*

In addition to the previous proposition, Zagoraiou and Antognini discussed optimal designs also for estimation of θ_1 for the stationary Ornstein-Uhlenbeck process with the ratio $\frac{\sigma^2}{2\theta_1}$ known. We should, however, give some points concerning their contribution. Firstly, the assumption $\frac{\sigma^2}{2\theta_1} = 1$ significantly influenced the Fisher information matrix as it

is not just a constant multiple of the Fisher information, and thus we cannot leave out the part of the Fisher information corresponding to the parameter σ^2. Secondly, the assumption $\frac{\sigma^2}{2\theta_1} = 1$ is very binding in the sense that θ_1 inherits unpleasant properties connected with the discontinuity of the part of Fisher information matrix corresponding to σ^2 (We discuss this phenomenon in Part III), which leads to the following conclusion:

Proposition 46 ([110]). *The diagonal entry of the Fisher information matrix corresponding to the parameter θ_1 is Schur convex and thus does not attain its maximum.*

The previous proposition does not necessarily imply the non-existence of an optimal design for a given information function Φ.

4.4 Brownian motions

A Brownian motion is any process of the form $\{X(t)\}_{t\geq 0} = \{A_{\boldsymbol{\theta}}(t) + \sigma W(t)\}_{t\geq 0}$. Such process is a solution to the stochastic differential equation

$$\mathrm{d}X(t) = a_{\boldsymbol{\theta}}(t)\mathrm{d}t + \sigma \mathrm{d}W(t),$$

where $a(t)$ is the derivative of $A(t)$.

Asymptotically optimal designs for Brownian motions were to a certain extent studied by Sacks and Ylvisaker [92].

For some specific Brownian motions, Harman and Štulajter [40] and Harman [37] derived exactly optimal designs.

Proposition 47 ([40]). *Let the observations follow $\{\theta_1 + \theta_2 t + \theta_3 t^2 + \sigma W(t)\}$, the number of observations $n \geq 0$, $T_* \leq t_1 < t_n \leq T^*$ and let $\delta_n \equiv [3 - (9 - 24n + 12n^2)^{1/2}]/(2n^2 - 4n) + 1$. Further define the set*

$$\mathcal{A}_n \equiv \{(t_1, t_n)^\mathsf{T} \in [T_*, T^*]^2 \mid t_1 \leq t_n - \delta_n(t_n - T_*), t_n \geq t_1 + \delta_n(T^* - t_1)\}.$$

Then, for any design $\boldsymbol{\tau} \in \mathfrak{T}_n$, there exists an equidistant design $\tilde{\boldsymbol{\tau}}$ with $(\tilde{t}_1, \tilde{t}_n)^\mathsf{T} \in \mathcal{A}_n$ such that $\mathscr{I}(\boldsymbol{\tau}) \preceq_{\mathrm{L}} \mathscr{I}(\tilde{\boldsymbol{\tau}})$. Moreover, equidistant design with $t_1 = T_*$ and $t_n = T^*$ is D-, A-, $\boldsymbol{e}_1$-, $\boldsymbol{e}_2$- and $\boldsymbol{e}_3$-optimal.

Proposition 48 ([40]). *Let the observations follow $\{\theta_1 + \theta_2 t + \theta_3 t^2 + \sigma W(t)\}$ and $T_* \leq t_1 < t_n \leq T^*$ and $\gamma_n = (n^2 - 2n)/(n^2 - 2n + 1)$. Then $\gamma_n \mathscr{I}_\infty(t_1, t_n) \preceq_{\mathrm{L}} \mathscr{I}(t_1, t_n)$, where $\mathscr{I}_\infty(t_1, t_n)$ is the asymptotic Fisher information matrix on $[t_1, t_n]$ and $\mathscr{I}(t_1, t_n)$ is the Fisher information matrix of the equidistant design.*

Proposition 49 ([37]). *Suppose we can observe a process $\{\boldsymbol{f}^\mathsf{T}(t)\boldsymbol{\theta} + \sigma W(t)\}_{t\geq 0}$.*

i) If $\boldsymbol{f}(t) = (1, t, \sqrt{t})^\mathsf{T}$ (or any of its submodels), then the optimal n-point design is generated by a progression with $t_i^* = (T_*)^{(n-i)/(n-1)}(T^*)^{(i-1)/(n-1)}$, $i = 1, \ldots, n$.

ii) If $\boldsymbol{f}(t) = (1, \mathrm{e}^{-\lambda t}, \mathrm{e}^{\lambda t})^\mathsf{T}$ (or any of its submodels), then the optimal n-point design is generated by a progression with $t_i^* = \frac{n-i}{n-1}T_* + \frac{i-1}{n-1}T^*$, $i = 1, \ldots, n$.

iii) If $\boldsymbol{f}(t) = (1, t, 1/t)^\mathsf{T}$ (or any of its submodels), then the optimal n-point design is generated by a progression with $t_i^* = (n-1)/(\frac{n-i}{T_*} + \frac{i-1}{T^*})$, $i = 1, \ldots, n$. The same hold true also in the instance when the error process is the Brownian bridge.

4.5 Compartmental models with stochastic trajectories

Jacquez [48] described compartmental model as "a system which is made up of a finite number of macroscopic subsystems, called compartments or pools, each of which is homogeneous and well mixed, and the compartments interact by exchanging materials. There may be inputs from the environment into one or more compartments, and there may be outputs (excretion) from one or more compartments into the environment." Compartmental models are strongly related to applications in medicine (physiology, pharmacokinetics), biology (ecosystem ecology), agriculture (fertilizer response), population movements or epidemics; see, Seber and Wild [97].

We usually define compartmental models by using the system of differential equations. In this section we focus on linear systems of differential equations of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}(t) = \mathbf{A}(\boldsymbol{\theta})\boldsymbol{X}(t) + \boldsymbol{b}(t), \tag{4.3}$$

where the vector $\boldsymbol{X}(t)$ represents the amount of material in each compartment at the time t, $\mathbf{A}$ is the matrix of transfer rates and $\boldsymbol{b}(t)$ is the vector of material inputs and outputs at the time t.

The deterministic dynamic system (4.3) yields a multiresponse nonlinear regression model

$$\boldsymbol{X}(t) = \mathcal{E}_{\boldsymbol{\theta}}[X(t)] + \boldsymbol{\varepsilon}(t)$$

where $\boldsymbol{\varepsilon}(t_i)$, $i = 1, \ldots$, are independent zero-mean random vectors with a given covariance matrix. We remark that for a given sampling design $\boldsymbol{\tau}_n \in \overline{\mathfrak{T}}_n$ we obtain vector-valued observations $\boldsymbol{X}(t_1), \ldots, \boldsymbol{X}(t_n)$.

In applications, we often cannot observe the values of all components of $\boldsymbol{X}(t)$, but a subset of them (usually one), i.e., we may consider scalar observations of the form $Y(t) = \boldsymbol{h}^{\mathsf{T}}\boldsymbol{X}(t)$. Then, $\mathcal{E}_{\boldsymbol{\theta}}[Y(t)] = \boldsymbol{h}^{\mathsf{T}}\mathcal{E}_{\boldsymbol{\theta}}[\boldsymbol{X}(t)]$ and $\mathcal{V}[Y(t)] = \boldsymbol{h}^{\mathsf{T}}\mathcal{V}[\boldsymbol{X}(t)]\boldsymbol{h}$, and, under the assumption of Gaussian errors, we can use the standard formula for the Fisher information matrix, cf. Proposition 12 or [67].

Regarding the subject of this thesis, we discuss extensions of the deterministic dynamic system (4.3) to stochastic dynamic systems represented by (systems of) stochastic differential equations.

4.5.1 A naive two compartmental model with stochastic trajectories

A direct extension to account for intrinsic within-object variability, Anisimov, Fedorov and Leonov [3] proposed to model the responses of the compartmental models as follows:

$$\mathrm{d}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} -\theta_1 & 0 \\ \theta_1 & \theta_2 \end{pmatrix}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \sigma_1(t) & 0 \\ 0 & \sigma_2(t) \end{pmatrix}\mathrm{d}\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}, \tag{4.4}$$

where $X_1(0) = D$ is the initial dose, $X_2(0) = 0$, $\sigma_i(t)$'s are non-negative deterministic functions and $(W_1(t), W_2(t))^{\mathsf{T}}$ is a two-dimensional Wiener process. The variable subject to experimental examination is $X_2(t)$.

The expectation of the system (4.4), which follows from the deterministic part, is given by $\mathcal{E}[X_1(t)] = D\mathrm{e}^{-\theta_1 t}$ and $\mathcal{E}[X_2(t)] = \frac{\theta_1 D}{\theta_1 - \theta_2}(\mathrm{e}^{-\theta_2 t} - \mathrm{e}^{-\theta_1 t})$. Anisimov et al. [3] further derived the covariance structure for observations $\boldsymbol{X}_2(\boldsymbol{\tau}_n)$ denoted by $\mathbf{S}(\boldsymbol{\tau}_n)$, which is complicated and, thus, is omitted.

Evidently, searching for optimal designs for estimation of unknown parameters of equation (4.4) is, again, a non-convex optimisation problem. Anisimov et al. [3] in their paper

avoided this difficulty by considering a population dynamics problem, where the unknown parameters $\tilde{\boldsymbol{\theta}}$ in individual experiments are sampled randomly from a prescribed distribution corresponding to the population variability, usually $\mathcal{N}(\boldsymbol{\theta}, \mathbf{D})$, and observations are measured with some noise, say δ^2. If δ^2 is small enough, then the variance of the observations is approximately [33]

$$\boldsymbol{\Sigma}(\boldsymbol{\tau}_n) \approx \mathbf{S}(\boldsymbol{\tau}_n) + \delta^2\mathbf{I} + \frac{\partial \mathcal{E}[\boldsymbol{X}(\boldsymbol{\tau}_n)]}{\partial \boldsymbol{\theta}^\mathsf{T}}\mathbf{D}\frac{\partial \mathcal{E}^\mathsf{T}[\boldsymbol{X}(\boldsymbol{\tau}_n)]}{\partial \boldsymbol{\theta}},$$

and for an experiment performed on some individual $\mathscr{I}(\boldsymbol{\tau}_n, \boldsymbol{\theta}^*)$ based on the well-known formula of Mardia and Marshall [67]; see also Proposition 12.

The key here is the fact that Anisimov et al. assumed a set of design measures Ξ on the set of sampling designs $\overline{\mathfrak{T}}_n$. For a given design $\xi \in \Xi$, the Fisher information matrix for the population experiment is

$$\mathscr{I}_{\mathrm{pop}}(\boldsymbol{\tau}, \boldsymbol{\theta}^*) = \int_{\boldsymbol{\tau}_n \in \overline{\mathfrak{T}}_n} \mathscr{I}(\boldsymbol{\tau}_n, \boldsymbol{\theta}^*)\mathrm{d}\xi(\boldsymbol{\tau}_n).$$

That is, we design an experiments with uncorrelated errors.

We should note that the Fisher information matrices of individual patients in population $\mathscr{I}(\boldsymbol{\tau}_n, \boldsymbol{\theta}^*)$ are not generally rank one matrices, hence we cannot use the classical theory of experimental design, cf. Section 2.4. Nevertheless, computational aspects are covered by a number of publications and the reader is referred to Yu [109], where a survey on existing algorithms and proof of their monotonic convergence is given.

4.5.2 A two compartmental model with positive stochastic trajectories

The dynamic system (4.4) allows negative values, which might not satisfy theoretical assumptions in certain applications. For that reason, Fedorov, Leonov and Vasiliev [26] extended the system 4.4 by adding proportionality of the governed variable into autonomous diffusion term as follows:

$$\mathrm{d}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} -\theta_1 & 0 \\ \theta_1 & \theta_2 \end{pmatrix}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \sigma_1 X_1(t) & 0 \\ 0 & \sigma_2 X_2(t) \end{pmatrix}\mathrm{d}\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}. \tag{4.5}$$

Again, the expectation of (4.5) comes from the deterministic part of the dynamic system. The solution of the system of stochastic differential equations (4.5) is known in an explicit closed form. An important fact is that the solution is not a Gaussian process, although Fedorov et al. used the Fisher information matrix for a Gaussian model as a proxy for the true Fisher information matrix.

Part III

Contribution

Chapter 5

Experiments with nonstationary Ornstein-Uhlenbeck processes with time-dependent volatility

The model underlying this chapter is a nonautonomous nonstationary Ornstein-Uhlenbeck process, that is, an Itō process $\{X(t)\}_{t\geq 0}$ governed by a stochastic differential equation of the form

$$\begin{aligned} \mathrm{d}X(t) &= \kappa(\bar{X} - X(t))\mathrm{d}t + \sigma(t)\mathrm{d}W(t), \\ X_0 &\quad \text{unknown}, \end{aligned} \tag{5.1}$$

where the initial point X_0 and the asymptotic expectation $\bar{X}$ are unknown parameters, and $\kappa > 0$ is known mean-reversion speed, $\sigma(\cdot) : \langle 0, \infty) \mapsto (0, \infty)$ is (up to a constant multiple) known deterministic and semicontinuous volatility function. The observations are taken at given sampling times represented by a sampling design $\boldsymbol{\tau}_n \in \overline{\mathfrak{T}}_n$.

The purpose of this chapter, which is based on the candidate's paper [59], is to discuss qualitative aspects of optimal n-point designs for estimation of the parameters of the model (5.1). In this regard, we focus also on the ultimate efficiency of experimental designs.

The model (5.1) is motivated by the (autonomous) nonstationary Ornstein-Uhlenbeck process, which corresponds to the stochastic differential equation (5.1) with a constant volatility function $\sigma(t) \equiv \sigma$. The nonautonomous Ornstein-Uhlenbeck process has found many applications in different research fields such as physics or biology. In physics [62, 96], the nonstationary Ornstein-Uhlenbeck process is a noise relaxation process, which describes the velocity of a particle under the influence of a friction. If we consider a Hookean spring, then the mean-reversion speed κ is given by the ratio of the spring constant k and the friction coefficient γ, and the volatility $\sigma(t)$ is constantly equal to $\sqrt{2k_{\mathrm{B}}T/\gamma}$, where k_{B} is the Boltzmann constant and T is the temperature. In biology, the nonstationary Ornstein-Uhlenbeck process is often employed for modelling neuronal response [90]. The governed variable $X(t)$ expresses the voltage difference between the membrane and resting potentials at the trigger zone of the neuron, and we are interested in the initial and stationary difference in potentials. Here, the mean-reversion speed is the reciprocal of the membrane constant.

From the point of view of optimal design of experiments, nonstationary Ornstein-Uhlenbeck process and its stationary counterpart have been studied by Harman and Štulajter [39], who showed the optimality of equidistant sampling designs for estimation of the pa-

rameteres as well as for the prediction, and by Kiseľák and Stehlík [53] and Zagoraiou and Antognini [110], who proved that the equidistant sampling is optimal also for parameter estimation of the stationary Ornstein-Uhlenbeck process. We refer the reader to Chapter 4 for more details.

However, the situation is quite different if we expect, for example, that in the problem of the Hookean spring the temperature T is a function of the time (e.g., cooling or heating of the physical system), and we get $\sigma(t) = \sqrt{2k_{\mathrm{B}}T(t)/\gamma}$. Since the volatility is not constant, as it can be seen in the (autonomous) nonstationary Ornstein-Uhlenbeck process, we should use the model (5.1) instead.

5.1 Corresponding linear regression model

After applying the Itō's lemma to transformation $\mathrm{e}^{\kappa t}(x - \bar{X})$ we obtain the solution to the stochastic differential equation (5.1)

$$X(t) = \mathrm{e}^{-\kappa t}X_0 + (1 - \mathrm{e}^{-\kappa t})\bar{X} + \int_0^t \mathrm{e}^{-\kappa(t-\nu)}\sigma(\nu)\mathrm{d}W(\nu). \tag{5.2}$$

Thus, the observations of the process driven by (5.1) at the design points $t_1, \ldots, t_n$ satisfy the linear regression model

$$\boldsymbol{X}(\boldsymbol{\tau}) = (\mathbf{e}^{-\kappa\boldsymbol{\tau}})X_0 + (\mathbf{1}_n - \mathbf{e}^{-\kappa\boldsymbol{\tau}})\bar{X} + \boldsymbol{\varepsilon}(\boldsymbol{\tau}) = \mathbf{F}(\boldsymbol{\tau})\boldsymbol{\theta} + \boldsymbol{\varepsilon}(\boldsymbol{\tau}), \tag{5.3}$$

where $\mathbf{F}(\boldsymbol{\tau}) = (\mathbf{e}^{-\kappa\boldsymbol{\tau}}, \mathbf{1}_n - \mathbf{e}^{-\kappa\boldsymbol{\tau}})$ is the design matrix, $\boldsymbol{\theta} = (X_0, \bar{X})^{\mathsf{T}}$ is the vector of unknown parameters, and $\boldsymbol{\varepsilon}(\boldsymbol{\tau}) = (\varepsilon_{t_1}, \ldots, \varepsilon_{t_n})^{\mathsf{T}}$ is the vector of random errors such that

$$\mathcal{E}[\boldsymbol{\varepsilon}(\boldsymbol{\tau})] = \mathbf{0}_n \text{ and } \mathcal{V}[\boldsymbol{\varepsilon}(\boldsymbol{\tau})] = \boldsymbol{\Sigma}(\boldsymbol{\tau}). \tag{5.4}$$

We remark that the distribution of the vector $\boldsymbol{\varepsilon}(\boldsymbol{\tau})$ is Gaussian. In the sequel we will derive the variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\tau})$, which is crucial for computing the information matrix $\boldsymbol{M}(\boldsymbol{\tau})$, and show that $\boldsymbol{\Sigma}(\boldsymbol{\tau})$ is positive definite for any $\boldsymbol{\tau} \in \mathfrak{T}_n$.

We shall consider conditioning upon the value x_0 of $X(0)$. A basic rule for covariance gives that

$$\begin{aligned}\mathcal{C}[X(t), X(t+s) \mid X(0) = x_0] \\ = \mathcal{E}[X(t)X(t+s) \mid X(0) = x_0] - \mathcal{E}[X(t) \mid X(0) = x_0]\mathcal{E}[X(t+s) \mid X(0) = x_0].\end{aligned}$$

The expectations $\mathcal{E}[X(t) \mid X(0) = x_0]$ and $\mathcal{E}[X(t+s) \mid X(0) = x_0]$ are known (cf. (5.2)), henceforth we need to find $\mathcal{E}[X(t)X(t+s) \mid X(0) = x_0]$.

The key for computing $\mathcal{E}[X(t)X(t+s) \mid X(0) = x_0]$ is the transition kernel $p(x, s \mid y, t) = \frac{\mathrm{d}}{\mathrm{d}x}\Pr[X(t+s) < x \mid X(t) = y]$, $t, s \geq 0$, of the process $X(t)$ defined by the stochastic differential equation (5.1), which solves the well-known Kolmogorov's forward equation.

Since $X(t)$ is a Markov process, for all $\varpi \in (0, t)$,its transition density kernel satisfies

$$\Pr[X(t) < x \mid X(0) = x_0] = \int_{-\infty}^{x} \left(\int_{\mathbb{R}} p(z, t - \varpi \mid y, \varpi)p(y, \varpi \mid x_0, 0)\mathrm{d}y \right) \mathrm{d}z.$$

Consequently, using (5.2) we obtain

$$\begin{aligned}\mathcal{E}[X(t)X(t+s) \mid X(0)=x_0] &= \int_{\mathbb{R}} x_1 f(x_1,t \mid x_0,0)\left(\int_{\mathbb{R}} x_2 f(x_2,s \mid x_1,t)\mathrm{d}x_2\right)\mathrm{d}x_1 \\ &= \int_{\mathbb{R}} x\mathcal{E}[X(t+s) \mid X(t)=x]f(x,t \mid x_0,0)\mathrm{d}x \\ &= \mathrm{e}^{-\kappa s}\mathcal{E}[X^2(t) \mid X(0)=x_0] \\ &\quad +\bar{X}(1-\mathrm{e}^{-\kappa s})\mathcal{E}[X(t) \mid X_0=x_0] \\ &= \mathrm{e}^{-\kappa s}\mathcal{V}[X(t) \mid X_0=x_0]+\mathrm{e}^{-\kappa s}\mathcal{E}^2[X(t) \mid X(0)=x_0] \\ &\quad +\bar{X}(1-\mathrm{e}^{-\kappa s})\mathcal{E}[X(t) \mid X(0)=x_0],\end{aligned}$$

which implies that

$$\mathcal{C}[X(t),X(t+s) \mid X(0)=x_0]=\mathrm{e}^{-\kappa s}\mathcal{V}[X(t) \mid X_0=x_0]=\mathrm{e}^{-\kappa s}\mathcal{V}[X(t)]. \tag{5.5}$$

It can be shown (see [4]) that the variance of $X(t)$ governed by stochastic differential equation (5.1) follows the ordinary differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}[X(t)]=-2\kappa\mathcal{V}[X(t)]+\sigma^2(t)$ with the initial condition $\mathcal{V}[X(0)]=0$. Using standard methods of solving ordinary differential equations we obtain the explicit solution

$$\mathcal{V}[X(t)]=\mathrm{e}^{-2\kappa t}\int_0^t \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu. \tag{5.6}$$

The same result follows from the Itō isometry:

$$\mathcal{E}\left[\int_0^t f(\nu)\mathrm{d}W(\nu)\right]=0, \text{ and } \mathcal{E}\left[\left(\int_0^t f(\nu)\mathrm{d}W(\nu)\right)^2\right]=\mathcal{V}\left[\int_0^t f(\nu)\mathrm{d}W(\nu)\right]=\int_0^t f^2(\nu)\mathrm{d}\nu.$$

The relations (5.5) and (5.6) yield:

Lemma 1. *The ijth element, $i\le j$, of the variance-covariance matrix $\mathbf{\Sigma}(\boldsymbol{\tau})$ defined in* (5.4) *has the form*

$$\begin{aligned}\{\mathbf{\Sigma}(\boldsymbol{\tau})\}_{ij} &= u(t_i)v(t_j), \text{ where} \\ u(t) &= \mathrm{e}^{-\kappa t}\int_0^t \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu \text{ and} \\ v(t) &= \mathrm{e}^{-\kappa t}.\end{aligned} \tag{5.7}$$

The least squares estimator and Fisher information matrix assume the invertibility of the variance-covariance matrix $\mathbf{\Sigma}(\boldsymbol{\tau})$. The next lemma states that $\mathbf{\Sigma}(\boldsymbol{\tau})$ is positive definite.

Lemma 2. *The variance-covariance matrix $\mathbf{\Sigma}(\boldsymbol{\tau})$ given by* (5.7) *is positive definite for any $\boldsymbol{\tau}\in\mathfrak{T}_n$, $n\ge 2$.*

Proof. For a design $(t_1,t_2)^{\mathsf{T}}\in\mathfrak{T}_2$, i.e., $t_1<t_2$, we have $\{\mathbf{\Sigma}(t_1,t_2)\}_{11}=\mathcal{V}[X(t_1)]>0$ and $\det[\mathbf{\Sigma}(t_1,t_2)]=\mathcal{V}[X(t_1)]|\mathcal{V}[X(t_2)]-\mathrm{e}^{-2\kappa(t_2-t_1)}\mathcal{V}^2[X(t_1)]>0$ because

$$\begin{aligned}\mathcal{V}[X(t_2)] &= \mathrm{e}^{-2\kappa t_2}\int_0^{t_2}\mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu \\ &> \mathrm{e}^{-2\kappa(t_2-t_1)}\mathrm{e}^{-2\kappa t_1}\int_0^{t_1}\mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu=\mathrm{e}^{-2\kappa(t_2-t_1)}\mathcal{V}[X(t_1)].\end{aligned} \tag{5.8}$$

Now, assume that $\boldsymbol{\tau}_n = (t_1, \dots, t_n)^\mathsf{T} \in \mathfrak{T}_n$, $\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)$ be positive definite, and, without loss of generality, $\boldsymbol{\tau}_{n+1} = (\boldsymbol{\tau}_n^\mathsf{T}, t_{n+1})^\mathsf{T} \in \mathfrak{T}_{n+1}$. Then

$$\boldsymbol{\Sigma}(\boldsymbol{\tau}_{n+1}) = \begin{pmatrix} \boldsymbol{\Sigma}(\boldsymbol{\tau}_n) & \boldsymbol{s} \\ \boldsymbol{s}^\mathsf{T} & \mathcal{V}[X(t_{n+1})] \end{pmatrix},$$

where $\boldsymbol{s} = (\mathrm{e}^{-\kappa(t_{n+1}-t_1)}\mathcal{V}[X(t_1)], \dots, \mathrm{e}^{-\kappa(t_{n+1}-t_n)}\mathcal{V}[X(t_n)])^\mathsf{T}$. Since $\boldsymbol{\Sigma}(\boldsymbol{\tau}_n)$ is nonsingular, the matrix $\boldsymbol{\Sigma}(\boldsymbol{\tau}_{n+1})$ is row-equivalent to

$$\begin{pmatrix} \boldsymbol{\Sigma}(\boldsymbol{\tau}_n) & \boldsymbol{s} \\ \mathbf{0}_n^\mathsf{T} & \mathcal{V}[X(t_{n+1})] - \boldsymbol{s}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{s} \end{pmatrix}.$$

The expression $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{s}$ is equal to the nth unit vector $(0, \dots, 0, 1)^\mathsf{T}$, cf. [39], hence we can write $\boldsymbol{s}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{s} = \mathrm{e}^{-\kappa(t_{n+1}-t_n}\mathcal{V}[X(t_n)]$. We can use the relation (5.8) to prove the positivity of $\mathcal{V}[X(t_{n+1})] - \boldsymbol{s}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}_n)\boldsymbol{s}$. □

5.2 Information matrix and optimal designs

The fact that $\boldsymbol{\Sigma}(\boldsymbol{\tau})$ has a product structure and the results of Harman and Štulajter [39] imply that for $\boldsymbol{\tau} \in \mathfrak{T}_n$

$$\mathscr{I}(\boldsymbol{\tau}) = \begin{pmatrix} \frac{\mathrm{e}^{-2\kappa t_1}}{\mathcal{V}[X(t_1)]} & \frac{\mathrm{e}^{-\kappa t_1}(1-\mathrm{e}^{-\kappa t_1})}{\mathcal{V}[X(t_1)]} \\ \frac{\mathrm{e}^{-\kappa t_1}(1-\mathrm{e}^{-\kappa t_1})}{\mathcal{V}[X(t_1)]} & \frac{(1-\mathrm{e}^{-\kappa t_1})^2}{\mathcal{V}[D(t_1)]} + S(\boldsymbol{\tau}) \end{pmatrix}, \tag{5.9}$$

where

$$S(\boldsymbol{\tau}) = \sum_{i=2}^{n} \frac{(\mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}})^2}{\mathrm{e}^{2\kappa t_i}\mathcal{V}[X(t_i)] - \mathrm{e}^{2\kappa t_{i-1}}\mathcal{V}[X(t_{i-1})]} = \sum_{i=2}^{n} \frac{(\mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}})^2}{\int_{t_{i-1}}^{t_i} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu}. \tag{5.10}$$

We remark that for $\boldsymbol{\tau} \in \mathfrak{T}_n$ the information matrix $\mathscr{I}(\boldsymbol{\tau})$ given in (5.9) is positive definite. Since the optimality criteria are usually continuous in $\mathscr{I}$ on the set of positive definite matrices and $\mathscr{I}(\boldsymbol{\tau})$ is continuous in $\boldsymbol{\tau}$ on $\mathfrak{T}_n$, the function $\Phi[\mathscr{I}(\boldsymbol{\tau})]$ is continuous in $\boldsymbol{\tau}$ on $\mathfrak{T}_n$.

Now, assume that $t_1^* \geq T_*$ is fixed. It follows from the Loewner isotonicity of optimality criteria that the design $\boldsymbol{\tau}^* \in \mathfrak{T}_n$ is optimal, if

$$S(\boldsymbol{\tau}^*) = \max_{\substack{t_2, \dots, t_n \\ t_1^* < t_2, t_n \leq T^* \\ t_{i-1} < t_i, i = 3, \dots, n}} S(t_1^*, t_2, \dots, t_n). \tag{5.11}$$

Therefore, once we have chosen the value of t_1^*, the other design points solve the optimization problem (5.11). This approach is very suitable for a numerical optimization: firstly, we select the value of t_1 and then we find the maximum of $S(\boldsymbol{\tau}^*)$ on $\mathfrak{T}_n$ with t_1 given, which is used for evaluation of the optimality criterion. In this way we can find the maximum of the optimality criterion through t_1. Our numerical experience shows that this approach gives more reliable results compared to a raw maximization of $\Phi[\mathscr{I}(\boldsymbol{\tau})]$ on the n-simplex $\mathfrak{T}_n$, even if we employ heuristic methods like simulated annealing or genetic algorithms.

The following lemma ensures that an optimal design is not degenerated into one point.

Lemma 3. *Let $n \geq 3$ and $\boldsymbol{\tau}_0 = t_n \mathbf{1}_n$. Then there exists $\boldsymbol{\tau}_1 = (t_1, \ldots, t_n)^\mathsf{T} \in \overline{\mathfrak{T}}_n$ such that $t_1 < t_n$ and $\mathscr{I}(\boldsymbol{\tau}_1) \succeq_\mathrm{L} \mathscr{I}(\boldsymbol{\tau}_0)$.*

Proof. Let $\boldsymbol{\tau}_1 = (t_1, \ldots, t_n)^\mathsf{T}$ with $t_1 < t_2 = \ldots = t_n$. To prove the statement of the lemma it is sufficient to show that the matrix

$$\mathscr{I}(\boldsymbol{\tau}_1) - \mathscr{I}(\boldsymbol{\tau}_0) = \begin{pmatrix} \frac{\mathrm{e}^{-2\kappa t_1}}{\mathcal{V}[X(t_1)]} - \frac{\mathrm{e}^{-2\kappa t_n}}{\mathcal{V}[X(t_n)]} & \frac{\mathrm{e}^{-\kappa t_1}(1-\mathrm{e}^{-\kappa t_1})}{\mathcal{V}[X(t_1)]} - \frac{\mathrm{e}^{-\kappa t_n}(1-\mathrm{e}^{-\kappa t_n})}{\mathcal{V}[X(t_n)]} \\ \frac{\mathrm{e}^{-\kappa t_1}(1-\mathrm{e}^{-\kappa t_1})}{\mathcal{V}[X(t_1)]} - \frac{\mathrm{e}^{-\kappa t_n}(1-\mathrm{e}^{-\kappa t_1})}{\mathcal{V}[X(t_n)]} & \frac{(1-\mathrm{e}^{-\kappa t_1})^2}{\mathcal{V}[X(t_1)]} - \frac{(1-\mathrm{e}^{-\kappa t_n})^2}{\mathcal{V}[X(t_n)]} + S(\boldsymbol{\tau}_1) \end{pmatrix}$$

is non-negative definite. From inequality (5.8) we have that $\mathcal{V}[X(t_n)] > \mathrm{e}^{-2\kappa(t_n - t_1)}\mathcal{V}[X(t_1)]$ for $t_1 < t_n$, which implies the positivity of $\{\mathscr{I}(\boldsymbol{\tau}_1) - \mathscr{I}(\boldsymbol{\tau}_0)\}_{11}$. Hence, we need to check whether $\det[\mathscr{I}(\boldsymbol{\tau}_1) - \mathscr{I}(\boldsymbol{\tau}_0)] \geq 0$. Notice that $S(\boldsymbol{\tau}_1) = \frac{(\mathrm{e}^{\kappa t_n} - \mathrm{e}^{\kappa t_1})^2}{\mathrm{e}^{2\kappa t_n}\mathcal{V}[X(t_n)] - \mathrm{e}^{2\kappa t_1}\mathcal{V}[X(t_1)]}$. Therefore,

$$\begin{aligned} \det[\mathscr{I}(\boldsymbol{\tau}_1) - \mathscr{I}(\boldsymbol{\tau}_0)] &= \frac{\mathrm{e}^{2\kappa t_n}\mathcal{V}[X(t_n)] - \mathrm{e}^{2\kappa t_1}\mathcal{V}[X(t_1)]}{\mathrm{e}^{2\kappa(t_1+t_n)}\mathcal{V}[X(t_1)]\mathcal{V}[X(t_n)]} S(\boldsymbol{\tau}_1) \\ &\quad - \frac{[\mathrm{e}^{-\kappa t_1}(1-\mathrm{e}^{-\kappa t_n}) - \mathrm{e}^{-\kappa t_n}(1-\mathrm{e}^{-\kappa t_1})]^2}{\mathcal{V}[X(t_1)]\mathcal{V}[X(t_n)]} \\ &= \frac{(\mathrm{e}^{\kappa t_n} - \mathrm{e}^{\kappa t_1})^2}{\mathrm{e}^{2\kappa(t_1+t_n)}\mathcal{V}[X(t_1)]\mathcal{V}[X(t_n)]} - \frac{(\mathrm{e}^{\kappa t_n} - \mathrm{e}^{\kappa t_1})^2}{\mathrm{e}^{2\kappa(t_1+t_n)}\mathcal{V}[X(t_1)]\mathcal{V}[X(t_n)]} = 0. \end{aligned}$$

□

Theorem 1. *Under the assumption of the model* (5.3) *with the covariance structure* (5.7)*, there always exists a (feasible) Φ-optimal n-point design $\boldsymbol{\tau}^*_{n,\Phi} \in \mathfrak{T}_n$.*

Proof. From Lemma 3 we get that $t_1 \neq t_n$, and the position of the design points $t_2, \ldots, t_n$ results from the optimization problem (5.11). Hence, we need to show that for any $t_i \in (t_{i-1}, t_{i+1})$, $i = 2, \ldots, n-1$,

$$\frac{(\mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}})^2}{\int_{t_{i-1}}^{t_i} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu} + \frac{(\mathrm{e}^{\kappa t_{i+1}} - \mathrm{e}^{\kappa t_i})^2}{\int_{t_i}^{t_{i+1}} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu} \geq \frac{(\mathrm{e}^{\kappa t_{i+1}} - \mathrm{e}^{\kappa t_{i-1}})^2}{\int_{t_{i-1}}^{t_{i+1}} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu}, \tag{5.12}$$

that is, we can place a design point t_i between any two design points t_{i-1} and t_{i+1} in a way that t_i is optimal and distinct from t_{i-1} and t_{i+1}. If we set $A = \mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}}$, $B = \mathrm{e}^{\kappa t_{i+1}} - \mathrm{e}^{\kappa t_i}$, $a = \int_{t_{i-1}}^{t_i} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu$ and $b = \int_{t_i}^{t_{i+1}} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu$, then the inequality (5.12) is equivalent to the statement

$$\forall_{A,B,a,b>0}\ \frac{A^2}{a} + \frac{B^2}{b} \geq \frac{(A+B)^2}{a+b}.$$

After some algebraic manipulation we obtain that $b^2A^2 - 2bAaB + a^2B^2 \geq 0$, which is true. Moreover, the inequality in (5.12) becomes an equality if and only if $t_i \in \{t_{i-1}, t_{i+1}\}$. □

The previous theorem states that the Fisher information matrix for the model (5.1) is continuous on $\overline{\mathfrak{T}}_n$.

In some applications it is natural to estimate the unknown parameters using the most recent observations. The derivative of $S(\boldsymbol{\tau})$ with respect to t_n

$$\frac{\partial S(\boldsymbol{\tau})}{\partial t_n} = [2\kappa - \sigma^2(t_n)\mathcal{U}(t_n, t_{n-1})]\mathcal{U}(t_n, t_{n-1}), \text{ where } \mathcal{U}(x, y) = \frac{\mathrm{e}^{2\kappa x} - \mathrm{e}^{\kappa(x+y)}}{\int_y^x \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu},$$

yields:

Theorem 2. *If $\sigma(t)$ is a nonincreasing function, then $t_n^* = T^*$ is optimal in the model* (5.3) *with the covariance structure* (5.7).

Proof. Clearly, the function $\mathcal{U}(x,y)$ is continuous and positive. For any $t_{n-1} < t_n \le T^*$ we have

$$\begin{aligned}\sigma^2(t_n)\mathcal{U}(t_n,t_{n-1}) &= \frac{\mathrm{e}^{2\kappa t_n} - \mathrm{e}^{\kappa(t_n+t_{n-1})}}{\int_{t_{n-1}}^{t_n} \mathrm{e}^{2\kappa\nu}\frac{\sigma^2(\nu)}{\sigma^2(t_n)}\mathrm{d}\nu} \le \frac{\mathrm{e}^{2\kappa t_n} - \mathrm{e}^{\kappa(t_n+t_{n-1})}}{\int_{t_{n-1}}^{t_n} \mathrm{e}^{2\kappa\nu}\mathrm{d}\nu} \\ &= 2\kappa\frac{\mathrm{e}^{2\kappa t_n} - \mathrm{e}^{\kappa(t_n+t_{n-1})}}{\mathrm{e}^{2\kappa t_n} - \mathrm{e}^{2\kappa t_{n-1}}} < 2\kappa,\end{aligned}$$

That is $\frac{\partial S(\boldsymbol{\tau})}{\partial t_n} > 0$ for any $\boldsymbol{\tau}$. □

Let us take into consideration that we can perform measurements at every point in the experimental domain $[T_*, T^*]$, and $\boldsymbol{\tau}_n = (t_1, \dots, t_n)^\mathsf{T}$ with $t_1 = T_*$ and $t_n = T^*$. From the Taylor series expansion of (5.10) and by setting $t_i - t_{i-1} = \Delta$, $i = 2, \dots, n$ we get that

$$\begin{aligned}S(\boldsymbol{\tau}_n) &= \kappa^2 \sum_{i=2}^{n} \frac{\left(\int_{t_{i-1}}^{t_{i-1}+\Delta} \mathrm{e}^{\kappa\nu}\mathrm{d}\nu\right)^2}{\int_{t_{i-1}}^{t_{i-1}+\Delta} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu} = \kappa^2 \sum_{i=2}^{n} \frac{\left(\mathrm{e}^{\kappa t_{i-1}}\Delta + \mathrm{o}(\Delta)\right)^2}{\mathrm{e}^{2\kappa t_{i-1}}\sigma^2(t_{i-1})\Delta + \mathrm{o}(\Delta)} \\ &= \kappa^2 \sum_{i=2}^{n} \frac{\mathrm{e}^{2\kappa t_{i-1}}\Delta^2 + \mathrm{o}(\Delta^2)}{\mathrm{e}^{2\kappa t_{i-1}}\sigma^2(t_{i-1})\Delta + \mathrm{o}(\Delta)} = \kappa^2 \sum_{i=2}^{n} \frac{\Delta + \mathrm{o}(\Delta)}{\sigma^2(t_{i-1}) + \mathrm{o}(\Delta)/\Delta}.\end{aligned}$$

Consequently,

$$S(\boldsymbol{\tau}_n) \to S_\infty(T_*, T^*) = \kappa^2 \int_{T_*}^{T^*} \frac{\mathrm{d}\nu}{\sigma^2(\nu)}, \text{ for } n \to \infty \text{ and } \Delta \to 0,$$

which, using the relation (5.12) and the technique in the proof of Lemma 3, leads to:

Theorem 3. *The information matrix given by the observation of the full path in the model* (5.1) *is*

$$\mathscr{I}_\infty(T_*, T^*) = \begin{pmatrix} \frac{\mathrm{e}^{-2\kappa T_*}}{\mathcal{V}[X(T_*)]} & \frac{\mathrm{e}^{-\kappa T_*}(1-\mathrm{e}^{-\kappa T_*})}{\mathcal{V}[X(T_*)]} \\ \frac{\mathrm{e}^{-\kappa T_*}(1-\mathrm{e}^{-\kappa T_*})}{\mathcal{V}[X(T_*)]} & \frac{(1-\mathrm{e}^{-\kappa T_*})^2}{\mathcal{V}[X(T_*)]} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} S_\infty(T_*, T^*). \tag{5.13}$$

Moreover, for any design $\boldsymbol{\tau} = (t_1, \dots, t_n)^\mathsf{T} \in \mathfrak{T}_n$ with $T_* \le t_1$ and $t_n \le T^*$ holds: ***i)*** *$S(\boldsymbol{\tau}) \le S_\infty(T_*, T^*)$, and* ***ii)*** *$\mathscr{I}(\boldsymbol{\tau}) \preceq_\mathrm{L} \mathscr{I}_\infty(t_1, t_n) \preceq_\mathrm{L} \mathscr{I}_\infty(T_*, T^*)$.*

The formula stated in (5.13) has an intuitive physical interpretation. In the Hookean spring problem with a time-dependent temperature, up to a constant multiple the function $\sigma(t)$ reflects the square root of the temperature. Thus the lower the temperature the higher the information about the unknown parameters. In the physical view, low temperature causes small fluctuations, so the measurements are more precise.

Another theoretical contribution of the formula (5.13) is the information contained in a subdomain interval. We shall explain this in detail. Let us consider that we perform one measurement at T_* and then we can observe the full trajectory of the process on a subdomain interval $(a, a+\Delta]$ of the fixed length Δ, where $a \in [T_*, T^* - \Delta]$. Then it is optimal to perform the measurements on such interval (determined by a), which maximizes $\int_a^{a+\Delta} \frac{\mathrm{d}\nu}{\sigma^2(\nu)}$. That is, the areas with low $\sigma(t)$ are more informative, and the measurements should be more concentrated in such areas. This effect is demonstrated on an example in the next section

We can use the asymptotic Fisher information matrix $\boldsymbol{M}_\infty(T_*, T^*)$ to evaluate the ultimate efficiency of designs.

5.3 Example

To give a simple demonstration of the previously presented results we will focus on D-optimal designs for

$$\mathrm{d}X(t) = \kappa(\bar{X} - X(t))\mathrm{d}t + \mathrm{e}^{-\lambda t}\mathrm{d}W(t), \tag{5.14}$$

where $\kappa > 0$ is the mean-reversion speed and $\lambda \in \mathbb{R}$ is a known constant. If λ is positive, then we have a system with exponentially decreasing temperature, and in the case of a negative value of λ the system is being heated.

Using the relations (5.6) and (5.9) we obtain that

$$\mathscr{I}_\lambda(\boldsymbol{\tau}) = \begin{pmatrix} \frac{2(\kappa-\lambda)}{\mathrm{e}^{2(\kappa-\lambda)t_1}-1} & \frac{2(\kappa-\lambda)(\mathrm{e}^{\kappa t_1}-1)}{\mathrm{e}^{2(\kappa-\lambda)t_1}-1} \\ \frac{2(\kappa-\lambda)(\mathrm{e}^{\kappa t_1}-1)}{\mathrm{e}^{2(\kappa-\lambda)t_1}-1} & \frac{2(\kappa-\lambda)(\mathrm{e}^{\kappa t_1}-1)^2}{\mathrm{e}^{2(\kappa-\lambda)t_1}-1} + S_\lambda(\boldsymbol{\tau}) \end{pmatrix},$$

where

$$S_\lambda(\boldsymbol{\tau}) = 2(\kappa-\lambda)\sum_{i=2}^{n} \frac{(\mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}})^2}{\mathrm{e}^{2(\kappa-\lambda)t_i} - \mathrm{e}^{2(\kappa-\lambda)t_{i-1}}}.$$

Obviously, for $\lambda \to \kappa$ we have

$$\mathscr{I}_\kappa(\boldsymbol{\tau}) = \begin{pmatrix} \frac{1}{t_1} & \frac{\mathrm{e}^{\kappa t_1}-1}{t_1} \\ \frac{\mathrm{e}^{\kappa t_1}-1}{t_1} & \frac{(\mathrm{e}^{\kappa t_1}-1)^2}{t_1} + S_\kappa(\boldsymbol{\tau}) \end{pmatrix}, \text{ with } S_\kappa(\boldsymbol{\tau}) = \sum_{i=2}^{n} \frac{(\mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}})^2}{t_i - t_{i-1}}.$$

Taking the D-optimality criterion $\Phi[\mathscr{I}] = \det^{1/2}(\mathscr{I})$ we get

$$\det{}^{1/2}(\mathscr{I}(\boldsymbol{\tau})) = \begin{cases} \left[\frac{2(\kappa-\lambda)}{\mathrm{e}^{2(\kappa-\lambda)t_1}-1} S_\lambda(\boldsymbol{\tau})\right]^{1/2}, & \lambda \neq \kappa \\ \left[\frac{1}{t_1} S_\kappa(\boldsymbol{\tau})\right]^{1/2}, & \lambda = \kappa \end{cases}.$$

We remark that for $\kappa = 2$ and $\lambda = 1$ the function $\mathcal{C}_\lambda(\boldsymbol{\tau})$ depends only on t_1 and t_n.

The ultimate Φ_D-efficiency of a design is given by

$$\mathrm{ueff}_{\Phi_D}(\boldsymbol{\tau}) = \begin{cases} \left[\frac{2\lambda}{\kappa^2} \cdot \frac{\mathrm{e}^{2(\kappa-\lambda)T_*}-1}{\mathrm{e}^{2(\kappa-\lambda)t_1}-1} \cdot \frac{S_\lambda(\boldsymbol{\tau})}{\mathrm{e}^{2\lambda T^*}-\mathrm{e}^{2\lambda T_*}}\right]^{1/2}, & \kappa \neq \lambda,\ \lambda \neq 0 \\ \left[\frac{1}{\kappa^2(T^*-T_*)} \cdot \frac{\mathrm{e}^{2(\kappa-\lambda)T_*}-1}{\mathrm{e}^{2(\kappa-\lambda)t_1}-1} S_\lambda(\boldsymbol{\tau})\right]^{1/2}, & \kappa \neq \lambda,\ \lambda = 0 \\ \left[\frac{2\lambda}{\kappa^2} \cdot \frac{T_*}{t_1} \cdot \frac{S_\kappa(\boldsymbol{\tau})}{\mathrm{e}^{2\lambda T^*}-\mathrm{e}^{2\lambda T_*}}\right]^{1/2}, & \kappa = \lambda \end{cases}.$$

In the sequel we give some numerical results for the model (5.14) with the mean-reversion speed $\kappa = 2$ and bounds for the experimental domain $T_* = 1$ and $T^* = 5$.

Figure 5.1 depicts D-optimal 5-point designs. More precisely, for a particular λ a horizontal cut gives optimal positions of the design points. For $\lambda = 0$ the design is equidistant, which is a known result already shown by Harman and Štulajter [40]. For $\lambda > 0$, in accord with Proposition 2, the position of the last design point is equal to $T^* = 5$. The explanation for the "jump" in optimal position of t_1 around $\lambda = 3.8$ can be found in flatness of the function $\varphi(t) = \max_{\boldsymbol{\tau} \in \mathfrak{T}_n(t)} \det^{1/2}(\boldsymbol{M}(\boldsymbol{\tau}))$, so the numerical optimization through t_1 is less accurate and more sensitive to roundoff errors. In both cases, $\lambda > 0$ and $\lambda < 0$, we can notice that if $|\lambda|$ is large, then the design points are more concentrated around T^* and T_*, respectively, where the fluctuations are smaller.

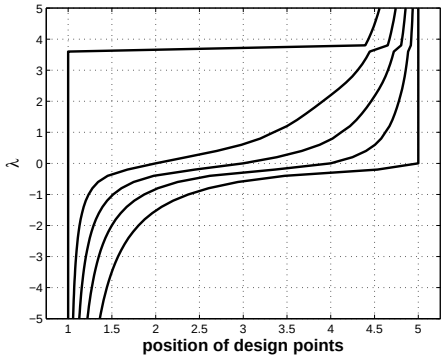


Figure 5.1. D-optimal 5-point designs for the model (5.14) with $\kappa = 2$, $T_* = 1$, $T^* = 5$ and different values of λ. For a given λ the horizontal cut gives the optimal position of design points.

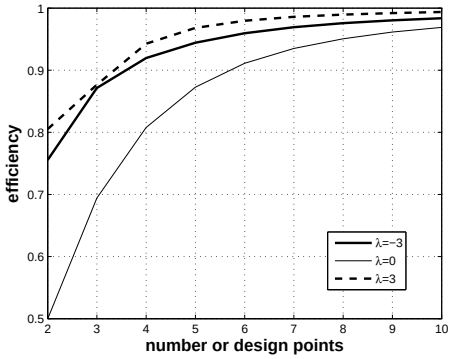


Figure 5.2. The relation between the size of a design sample n and efficiency of D-optimal n-point design with respect to the maximum possible information for the model (5.14) with $\kappa = 2$, $T_* = 1$, $T^* = 5$ and $\lambda = -3, 0, 3$.

In the previous section we noted that small sample designs can be quite efficient with respect to the maximum possible information, and the contributions of additional measurements to the information are not significant. This is illustrated in Figure 5.2, which displays the dependence of the efficiency of the D-optimal n-point designs on the size n of the design sample for the model (5.14).

Chapter 6

Ultimate efficiency of experimental designs for Ornstein-Uhlenbeck type processes

Suppose we can observe a univariate continuous-time process $\{X(t)\}_{t\geq 0}$ governed by a linear Itō stochastic differential equation of the form

$$\begin{aligned} \mathrm{d}X(t) &= [a_{\boldsymbol{\theta},\beta}(t) + b_{\boldsymbol{\theta},\beta}(t)X(t)]\mathrm{d}t + \sigma_\beta(t)\mathrm{d}W(t) \\ &= f_{\boldsymbol{\theta},\beta}(t, X(t))\mathrm{d}t + \sigma_\beta(t)\mathrm{d}W(t), \\ X(0) &= X_0 \in \mathbb{R} \text{ is fixed.} \end{aligned} \tag{6.1}$$

Our basic assumptions concerning the model are that the derivatives of the drift function $f_{\boldsymbol{\theta},\beta}(t,x)$ and volatility $\sigma_\beta(t)$ exist with respect to the unknown vector parameter $\boldsymbol{\vartheta} = (X_0, \boldsymbol{\theta}^\mathsf{T}, \beta)^\mathsf{T} \in \mathbb{R} \times \mathbb{R}^{m-2} \times \mathbb{R}$, that the functions $a_{\boldsymbol{\theta},\beta}(t)$, $\frac{\partial a_{\boldsymbol{\theta},\beta}(t)}{\partial \boldsymbol{\vartheta}}$, $b_{\boldsymbol{\theta},\beta}(t)$, $\frac{\partial b_{\boldsymbol{\theta},\beta}(t)}{\partial \boldsymbol{\vartheta}}$, $\sigma_\beta(t)$ and $\sigma^2_\beta(t)$ are integrable with respect to t on $[0, T^*]$, and that $\sigma_\beta(t)$ is positive almost everywhere with respect to the Lebesgue measure on the real axis. For convenience, we replace $a_{\boldsymbol{\theta},\beta}(t)$ by $a(t)$, $b_{\boldsymbol{\theta},\beta}(t)$ by $b(t)$, etc.

The choice of parametrisation in equation (6.1) reflects the different asymptotic properties of the maximum likelihood estimators. In Section 6.2 we show that, in contrast with the scalar parameter β, the vector parameter $\boldsymbol{\vartheta}_\mathrm{I} = (X_0, \boldsymbol{\theta}^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{m-1}$ is not consistently estimable. Notice that the initial value X_0 is absent in the governing equation. This consideration is not necessary for this paper, but its rejection might in some instances lead to "unexpected" results (see Theorem 4 and the discussion thereafter).

The aim of the presented chapter based on the candidate's paper [60] is two-fold. Firstly, we discuss the existence of optimal sampling designs, i.e., designs maximising the corresponding measure of information $\Phi[\mathscr{I}(\boldsymbol{\tau})]$. Although we do not determine optimal designs in this chapter, the question of their existence is essential as it gives a rationale for the optimisation of the experiment. Secondly, the presence of a correlation between observations of the process described by stochastic differential equation (6.1) makes searching for optimal designs computationally challenging. To circumvent this difficulty, in Section 6.3 we look at the experimental design problem from a different perspective by computing the ultimate efficiency of designs. This approach is, however, well-founded only if we focus on the parameter subvector $\boldsymbol{\vartheta}_\mathrm{I}$, which is not binding, because in the applications, $\boldsymbol{\vartheta}_\mathrm{I}$ usually represents the characteristics of the underlying process, while β is regarded as a nuisance parameter out of the perimeter of interest.

The governing equation (6.1) is motivated by a variety of problems studied in the literature; a number of them were already mentioned in the previous chapter. The first group of problems, which we already mentioned in the previous chapter, consists of modifications of the Ornstein-Uhlenbeck process. Uhlenbeck and Ornstein [104] proposed a simple model of particle velocity $X(t)$, which can be rewritten to the stochastic differential equation

$$\mathrm{d}X(t) = \left(\frac{\theta_1}{\theta_2} - \frac{\beta}{\theta_2}X(t)\right)\mathrm{d}t + \left(\frac{2\mathrm{B}T}{\beta}\right)^{1/2}\mathrm{d}W(t),\ X(0) = X_0 \text{ fixed}, \tag{6.2}$$

where θ_1 is the level of the external force, θ_2 is the mass, β is the friction coefficient, T is the temperature of the system, and B is the Boltzmann constant. We can find in the applications a more abstract arrangement of stochastic differential equation (6.2) given by

$$\mathrm{d}X(t) = \theta_1[\theta_2 - X(t)]\mathrm{d}t + \beta\mathrm{d}W(t),\ X(0) = X_0 \text{ fixed}, \tag{6.3}$$

where the mean-reversion speed θ_1, the asymptotic mean θ_2, and the diffusion coefficient β are the unknown parameters. Note that, compared to equation (6.2), the drift and diffusion part of the simpler variant (6.3) do not have any common parameters.

The second group of problems covered by model (6.1) is the family of Brownian motions of the form $\{A_{\boldsymbol{\theta}}(t) + \beta W(t)\}_{t\geq 0}$ with unknown parameters $\boldsymbol{\theta}$ and β, which coincides with a stochastic differential equation

$$\mathrm{d}X(t) = a_{\boldsymbol{\theta}}(t)\mathrm{d}t + \beta\mathrm{d}W(t),$$

where $a(t)$ is the derivative of $A(t)$ with respect to t.

In Section 6.5, as an example, we study the Gompertz model of tumour growth described by stochastic differential equation $\mathrm{d}Y(t) = [\theta_2 Y(t) - \theta_1 Y(t)\ln Y(t)]\mathrm{d}t + \beta Y(t)\mathrm{d}W(t)$. Although the governing equation is nonlinear, we show that the process belongs to a broader class of the so-called "Ornstein-Uhlenbeck type" processes defined in Section 6.4, to which the results of this paper are applicable.

Given the correlated observations, to write the Fisher information matrix in a suitable form is usually a crucial step. For the purpose of this paper we find the following lemma useful:

Lemma 4. *Let $\{X(t)\}_{t\geq 0}$ with $X(0)$ fixed be a $\boldsymbol{\vartheta}$-parametrised continuous-time Markov process with a Gaussian transition density kernel. Then for any $\boldsymbol{\tau} \in \overline{\mathfrak{T}}_n$, the Fisher information matrix for $\boldsymbol{X}(\boldsymbol{\tau})$ takes the form*

$$\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) = \mathscr{I}_{X(t_1)|X(0)}(\boldsymbol{\vartheta}^*) + \sum_{i=2}^{n} \mathcal{E}_{X(t_{i-1})}\left[\mathscr{I}_{X(t_i)|X(t_{i-1})}(\boldsymbol{\vartheta}^*)\right],$$

where $\mathscr{I}_{X(t_i)|X(t_{i-1})}(\boldsymbol{\vartheta}^*)$ is the Fisher information matrix for $X(t_i)$ conditioned on the value of $X(t_{i-1})$, and $\mathcal{E}_{X(t_{i-1})}[\cdot]$ is the expectation with respect to $X(t_{i-1})$.

Proof. We denote $t_0 = 0$ and consider the derivatives to be evaluated at $\boldsymbol{\vartheta}^*$. Let $p(x, s \mid y, t) = \frac{\mathrm{d}}{\mathrm{d}x}\Pr[X(t+s) < x \mid X(t) = y]$ be the Gaussian transition density kernel of the

process $\{X(t)\}_{t\geq 0}$, where $p(x,0 \mid y,t) = \delta(x-y)$, and $\delta(\cdot)$ is the Dirac delta function. Then

$$\begin{aligned}
\mathscr{I}(\boldsymbol{\tau},\boldsymbol{\vartheta}^*) &= \mathcal{E}_{\boldsymbol{X}(\boldsymbol{\tau})}\left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\mathsf{T}}\ln\left(\prod_{i=1}^{n} p(x_i, t_i - t_{i-1} \mid x_{i-1}, t_{i-1})\right)\right] \\
&= \int_{x_1\in\mathbb{R}}\cdots\int_{x_n\in\mathbb{R}}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\mathsf{T}}\ln p(x_i, t_i - t_{i-1} \mid x_{i-1}, t_{i-1})\right) \\
&\quad \times\prod_{k=1}^{n} p(x_k, t_k - t_{k-1} \mid x_{k-1}, t_{k-1})\mathrm{d}x_1\ldots\mathrm{d}x_n \\
&= \sum_{i=1}^{n}\int_{x_{i-1}\in\mathbb{R}}\int_{x_i\in\mathbb{R}}\left(\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\mathsf{T}}\ln p(x_i, t_i - t_{i-1} \mid x_{i-1}, t_{i-1})\right) \\
&\quad \times p(x_i, t_i - t_{i-1} \mid x_{i-1}, t_{i-1})p(x_{i-1}, t_{i-1} \mid X(0), 0)\mathrm{d}x_{i-1}\mathrm{d}x_i \\
&= \sum_{i=1}^{n}\int_{x_{i-1}\in\mathbb{R}}\mathscr{I}_{X(t_i)|X(t_{i-1})}(\boldsymbol{\vartheta}^*)p(x_{i-1}, t_{i-1} \mid X(0), 0)\mathrm{d}x_{i-1}.
\end{aligned}$$

□

Lemma 4 enables us to express the Fisher information matrix of Gaussian Markov processes in a practical form, which is also suitable for the asymptotic analysis in Sections 6.2 and 6.3. In addition, we note that the validity of Lemma 4 is not restricted only to univariate Gaussian Markov processes, but can be applied to any (multivariate) Markov process with a transition density satisfying classical regularity conditions, which the reader can find, for instance, in [61, Lem5.3, p116] or in Section 2.1.

6.1 Nonlinear regression model for observations

To assess the amount of information that an experimental design yields, we need to understand the mutual relations between individual observations $\boldsymbol{X}(\boldsymbol{\tau}) = (X(t_1),\ldots,X(t_n))^\mathsf{T}$. In this Section we formulate the model for $\boldsymbol{X}(\boldsymbol{\tau})$ in terms of Gaussian nonlinear regression models.

It is a well-known fact that by applying Itō's lemma (see Theorem 6 by Itō [47]) to the transformation $\mathrm{e}^{-B(t)}X(t)$, where $B(t)$ is an arbitrary antiderivative of $b(t)$, we can write $X(t) \mid X(t_0) = \mathcal{E}[X(t) \mid X(t_0)] + \varepsilon(t) \mid X(t_0)$ for all $t \geq t_0 \geq 0$. Here

$$\mathcal{E}_{\boldsymbol{\theta},\beta}[X(t) \mid X(t_0)] = \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(t_0)}X(t_0) + \int_{t_0}^{t}\mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(\nu)}a_{\boldsymbol{\theta},\beta}(\nu)\mathrm{d}\nu \tag{6.4}$$

is the expectation of the process $\{X(t) \mid X(t_0)\}_{t\geq t_0}$ at the time t, and

$$\varepsilon(t) \mid X(t_0) = \int_{t_0}^{t}\mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(\nu)}\sigma_\beta(\nu)\mathrm{d}W(\nu)$$

is a zero-mean Gaussian random variable; see, e.g., [35]. (Notice that the assumption $t \geq t_0$ is only for convenience and that the previous formulas also hold true for $t \leq t_0$.) The expression for the variance of $\{X(t) \mid X(t_0)\}_{t\geq t_0}$, which comes from Itō's isometry (see, e.g., Lemma 3.1.5 by [73]), is given by

$$\mathcal{V}_{\boldsymbol{\theta},\beta}[X(t) \mid X(t_0)] = \int_{t_0}^{t}\mathrm{e}^{2[B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(\nu)]}\sigma_\beta^2(\nu)\mathrm{d}\nu. \tag{6.5}$$

The following lemma completes an insight into second central moments:

Lemma 5. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by stochastic differential equation* (6.1)*. Then for all t_1 and t_2, $t_2 \geq t_1 \geq 0$,*

$$\begin{aligned} \mathcal{C}_{\boldsymbol{\theta},\beta}[X(t_1), X(t_2)] &= u(t_1)v(t_2), \text{ where} \\ u_{\boldsymbol{\theta},\beta}(t) &= \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)} \int_0^t \mathrm{e}^{-2B_{\boldsymbol{\theta},\beta}(\nu)} \sigma^2_\beta(\nu) \mathrm{d}\nu, \\ v_{\boldsymbol{\theta},\beta}(t) &= \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)}. \end{aligned}$$

Proof. We derive an ordinary differential equation for the covariance function. Since the process $\{X(t)\}_{t>0}$ is Gaussian, the usual regularity conditions for the interchange of the order of differentiation and integration are satisfied. Taking the expectation of the governing equation (6.1) yields $\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{E}[X(t+s)] = \mathcal{E}[f(t+s, X(t+s))]$, where we recall that $f(t,x) = a(t) + b(t)x$. Consequently, for all $t, s \geq 0$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s}\mathcal{E}[X(t)X(t+s)] &= \mathcal{E}\left[X(t)\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{E}[X(t+s) \mid X(t)]\right] = \mathcal{E}[X(t)f(t+s, X(t+s))], \\ \mathcal{E}[X(t)]\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{E}[X(t+s)] &= \mathcal{E}[X(t)]\mathcal{E}[f(t+s, X(t+s))]. \end{aligned}$$

The basic rules for covariance give an ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{C}[X(t), X(t+s)] = \mathcal{C}[X(t), f(t+s, X(t+s))] = b(t+s)\mathcal{C}[X(t), X(t+s)]$$

with the initial value $\mathcal{C}[X(t), X(t+s)] = \mathcal{V}[X(t)]$ at $s = 0$. The use of standard methods for solving ordinary differential equations and subsequent setting $t_1 = t$ and $t_2 = t + s$ entails that for any t_1 and t_2, $t_2 \geq t_1 \geq 0$, $\mathcal{C}[X(t_1), X(t_2)] = \mathrm{e}^{B(t_2)-B(t_1)}\mathcal{V}[X(t_1)]$. By setting the expression for the variance (6.5) to the formula obtained for the covariance, we get the statement of the lemma. □

The expectation of the underlying model (6.4) and Lemma 5 enable us to formulate the design problem for stochastic differential equation (6.1) in terms of the nonlinear regression

$$\forall_{\boldsymbol{\tau}\in\overline{\mathfrak{T}}_n}\ \boldsymbol{X}(\boldsymbol{\tau}) \sim \mathcal{N}\Big(\mathcal{E}_{X_0,\boldsymbol{\theta},\beta}[\boldsymbol{X}(\boldsymbol{\tau})], \mathcal{V}_{\boldsymbol{\theta},\beta}[\boldsymbol{X}(\boldsymbol{\tau})]\Big) \tag{6.6}$$

with $\{\mathcal{V}[\boldsymbol{X}(\boldsymbol{\tau})]\}_{ij} = u(t_i)v(t_j)$ for $1 \leq i \leq j \leq n$. We should highlight that the vector of the expected values and the variance-covariance matrix may have common parameters.

Note that $u(t)/v(t)$ is a strictly increasing function, and so it follows from the conditions (4.1) and (4.2) by Sacks and Ylvisaker[92] that for any design $\boldsymbol{\tau} \in \mathfrak{T}_n$, the matrix $\mathcal{V}[\boldsymbol{X}(\boldsymbol{\tau})]$ is positive definite – a desirable property that is not satisfied for designs with replicated sampling times. Nevertheless, by performing experiments under designs without replicated sampling times, which form a subset of $\overline{\mathfrak{T}}_n$, the amount of information does not decrease:

Lemma 6. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by stochastic differential equation* (6.1) *and $\|\cdot\|$ be a metric on $\mathbb{R}^n$. For any design $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n$ and $\delta > 0$, a design $\boldsymbol{\tau} \in \mathfrak{T}_n$ exists such that $\|\boldsymbol{\tau}_0 - \boldsymbol{\tau}\| < \delta$ and $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) \succeq_{\mathrm{L}} \mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\vartheta}^*)$. Specifically, if the design $\boldsymbol{\tau}_0$ belongs to the boundary set $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$, then $\boldsymbol{\tau}_0$ is dominated by any design $\boldsymbol{\tau} \in \mathfrak{T}_n$ with $t_1 = \{\boldsymbol{\tau}_0\}_1$ and $t_i = \{\boldsymbol{\tau}_0\}_i$ if $\{\boldsymbol{\tau}_0\}_i > \{\boldsymbol{\tau}_0\}_{i-1}$.*

Proof. If $\boldsymbol{\tau}_0 \in \mathfrak{T}_n$ then $\boldsymbol{\tau} = \boldsymbol{\tau}_0$. Assume that $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$, i.e., there exists $i \in \{2, \ldots, n\}$ such that $t_{i-1} = t_i$. Equations (6.4) and (6.5) yield $X(t) \mid X(t) = X(t)$ almost surely, which gives no information about any unknown parameters, so $\mathcal{E}_{X(t_i)}[\mathscr{I}_{X(t_i)|X(t_i)}(\boldsymbol{\vartheta}^*)] = \mathbf{0}_{m\times m}$.

As a consequence of Lemma 4, by leaving t_i out of the experimental design, the amount of information does not change. We repeat this procedure until we obtain a design $\boldsymbol{\tau}_1 \in \mathfrak{T}_{n_1,\mathcal{D}}$, $n_1 < n$, for which we have $\mathscr{I}(\boldsymbol{\tau}_1, \boldsymbol{\vartheta}^*) = \mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\vartheta}^*)$. For $\boldsymbol{\tau}$ we take any design from $\mathfrak{T}_n$ with arbitrary n_1 components being given by $\boldsymbol{\tau}_1$, for which $\|\boldsymbol{\tau}_0 - \boldsymbol{\tau}\| < \delta$. Let $\boldsymbol{\tau}_2$ represent those sampling times of $\boldsymbol{\tau}$ that are distinct from $\boldsymbol{\tau}_1$. Analogously to Lemma 4, we can show that $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) = \mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_1)|X(0)}(\boldsymbol{\vartheta}^*) + \mathcal{E}_{\boldsymbol{X}(\boldsymbol{\tau}_1)}[\mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_2)|\boldsymbol{X}(\boldsymbol{\tau}_1)}(\boldsymbol{\vartheta}^*)] = \mathscr{I}(\boldsymbol{\tau}_1, \boldsymbol{\vartheta}^*) + \mathcal{E}_{\boldsymbol{X}(\boldsymbol{\tau}_1)}[\mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_2)|\boldsymbol{X}(\boldsymbol{\tau}_1)}(\boldsymbol{\vartheta}^*)]$, which implies

$$\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) - \mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\vartheta}^*) = \mathcal{E}_{\boldsymbol{X}(\boldsymbol{\tau}_1)}[\mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_2)|\boldsymbol{X}(\boldsymbol{\tau}_1)}(\boldsymbol{\vartheta}^*)] \succeq_{\mathrm{L}} \mathbf{0}_{m\times m}.$$

□

The key elements in the proof of Lemma 6 are the Markov property of $\{X(t)\}_{t\geq 0}$ and the fact that $\lim_{\Delta\to 0} \Pr[X(t+\Delta) = X(t)] = 1$. Analogously to Lemma 4, we can extend the statement of Lemma 6 to a wider class of problems.

Under the regression model (6.6), for any design $\boldsymbol{\tau} \in \mathfrak{T}_n$, the Fisher information matrix takes the form

$$\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) = \begin{pmatrix} \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_1\boldsymbol{\vartheta}_1} & \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_1\beta} \\ \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\beta\boldsymbol{\vartheta}_1} & \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\beta\beta} \end{pmatrix},$$

where

$$\{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2} = \left(\frac{\partial \mathcal{E}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial \boldsymbol{\alpha}_1} \mathcal{V}^{-1}[\boldsymbol{X}(\boldsymbol{\tau})] \frac{\partial \mathcal{E}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial \boldsymbol{\alpha}_2^{\mathsf{T}}} \right) \Bigg|_{\boldsymbol{\vartheta}^*} + \frac{1}{2} \mathrm{tr} \left\{ \mathcal{V}^{-1}[\boldsymbol{X}(\boldsymbol{\tau})] \frac{\partial \mathcal{V}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial \boldsymbol{\alpha}_1} \mathcal{V}^{-1}[\boldsymbol{X}(\boldsymbol{\tau})] \frac{\partial \mathcal{V}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial \boldsymbol{\alpha}_2^{\mathsf{T}}} \right\} \Bigg|_{\boldsymbol{\vartheta}^*} \tag{6.7}$$

and $\boldsymbol{\vartheta}^*$ is a guess at the true value of the unknown vector parameter $\boldsymbol{\vartheta}$; see [67].

Theorem 4. *If the initial value X_0 of stochastic differential equation* (6.1) *is the only unknown parameter, then it is optimal to take $t_1 = T_*$ regardless of the number of design points. The variance of the corresponding maximum likelihood estimate is*

$$\mathcal{V}[\hat{X}_0] = \int_0^{T_*} \mathrm{e}^{2[B(0)-B(\nu)]} \sigma^2(\nu) \mathrm{d}\nu.$$

Proof. Since $\frac{\partial \mathcal{E}[X(t)]}{\partial X_0} = \mathrm{e}^{B(t)-B(0)} = v(t)\mathrm{e}^{-B(0)}$ and $\frac{\partial \mathcal{V}[X(t)]}{\partial X_0} = 0$, the relation (4.4) in the paper of Sacks and Ylvisaker [92] and Lemma 5 imply that for any design $\boldsymbol{\tau} \in \mathfrak{T}_n$, the information about X_0 is given by

$$\mathscr{I}(\boldsymbol{\tau}, X_0^*) = \mathcal{V}^{-1}[\hat{X}_0] = \left(\int_0^{t_1} \mathrm{e}^{2[B(0)-B(\nu)]} \sigma^2(\nu) \mathrm{d}\nu \right)^{-1}.$$

That is, the information is determined by the first sampling time t_1. Since the function $\mathrm{e}^{2[B(0)-B(t)]}\sigma^2(t)$ is positive for almost all t with respect to the Lebesgue measure on the real axis, $\mathscr{I}(\boldsymbol{\tau}, X_0^*)$ attains its maximum at $t_1 = T_*$. That is, the information is determined by the first sampling time t_1. Since the function $\mathrm{e}^{2[B(0)-B(t)]}\sigma^2(t)$ is positive for almost all t with respect to the Lebesgue measure on the real axis, $\mathscr{I}(\boldsymbol{\tau}, X_0^*)$ attains its maximum at $t_1 = T_*$. □

The statement of Theorem 4 might evoke the impression that the information about the initial state of the processes governed by a linear stochastic differential equation is concentrated in the first observation, but this is not true in general:

Example 1. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by equation*

$$\mathrm{d}X(t) = \frac{1}{2}\theta\left(X_0 - X(t)\right)\mathrm{d}t + \mathrm{e}^{-\beta t/2}\mathrm{d}W(t),\ X(0) = X_0, \tag{6.8}$$

for which we can perform only one observation. In this setup the values θ and β are known (and positive), while X_0 is an unknown parameter. Evidently, the model (6.8) *violates the assumption of the absence of the initial value in the governing equation. Here, the mean value is equal to X_0 for all $t \geq 0$, and the Fisher information obtained from the observation performed at the time $t \in \mathcal{D}$ attains the value*

$$\mathscr{I}(t,\theta,\beta) = \begin{cases} \dfrac{\theta-\beta}{\mathrm{e}^{-\beta t} - \mathrm{e}^{-\theta t}}, & \theta \neq \beta, \\ \dfrac{1}{\mathrm{e}^{-\theta t} t}, & \theta = \beta. \end{cases}$$

Since $\mathscr{I}(t,\theta,\beta)$ is a convex and continuous function, its maximum is allocated at the boundary of the experimental domain $\mathcal{D}$. For $\theta \neq \beta$ we can find that the sampling time minimising the Fisher information is given by $t_{\min\mathscr{I}} = (\ln\theta - \ln\beta)/(\theta-\beta)$ and $t_{\min\mathscr{I}} = 1/\theta$ for $\beta \to \theta$. Consequently, if the bounds of the experimental domain satisfy the inequality $T_* < t_{\min\mathscr{I}} < T^*$ with $\mathscr{I}(T_*,\theta,\beta) < \mathscr{I}(T^*,\theta,\beta)$ or the bounds satisfy $t_{\min\mathscr{I}} \leq T_* < T^*$, then it is optimal to observe the process as late as possible.

6.2 Existence of locally optimal designs

In this section we turn to the question of the existence of locally optimal sampling designs, which is crucial for a further (typically numerical) optimisation of the experiment. We say, as usual, that an optimal design exists if the measure of information achieves its maximum on the set of competing designs. Due to Lemma 6, the set of competing designs is reduced to the non-compact set $\mathfrak{T}_n$, and thus the maximum of $\Phi[\mathscr{I}(\boldsymbol{\tau},\boldsymbol{\vartheta}^*)]$ might not be achieved:

Example 2. *Suppose we can perform n observations of a process $\{X(t)\}_{t\geq 0}$ governed by a stochastic differential equation*

$$\mathrm{d}X(t) = (\beta+t)^{-1/2}\mathrm{d}W(t),\ X(0) = 0, \tag{6.9}$$

where $\beta \in (0,\infty)$ is an unknown parameter, and β^* is a prior guess at the true value of β. For a design $\boldsymbol{\tau} \in \mathfrak{T}_n$ Lemma 4 yields the Fisher information matrix

$$\mathscr{I}(\boldsymbol{\tau},\beta^*) = \left(\frac{\partial}{\partial\beta}\ln\mathcal{V}[X(t_1)]\right)^2\Bigg|_{\beta^*} + \sum_{i=2}^{n}\left(\frac{\partial}{\partial\beta}\ln\mathcal{V}[X(t_i)\mid X(t_{i-1})]\right)^2\Bigg|_{\beta^*}.$$

Using the expression for the variance (6.5)*, we obtain $(\frac{\partial}{\partial\beta}\ln\mathcal{V}[X(t_i)\mid X(t_{i-1})])^2 = [(\beta+t_i)^{-1} - (\beta+t_{i-1})^{-1}]^2[\ln(\beta+t_i)-\ln(\beta+t_{i-1})]^{-2}$, which is a function decreasing in t_i for any $\beta > 0$, so the amount of information increases as $t_i \to t_{i-1}$. Moreover, $\lim_{t_i\to t_{i-1}}(\frac{\partial}{\partial\beta}\ln\mathcal{V}[X(t_i)\mid X(t_{i-1})])^2 = (\beta+t_{i-1})^{-2}$; hence the optimum is attained for $t_1 \to T_*$ and $t_i \to t_{i-1}$ for all $i = 2,\ldots,n$ with the "optimal" value*

$$\lim_{t_1\to T_*,\ t_i\to t_{i-1}\ \forall_{i\geq 2}} \mathscr{I}(\boldsymbol{\tau},\beta^*) = \frac{[(\beta^*+T_*)^{-1} - \beta^{*-1}]^2}{[\ln(\beta^*+T_*) - \ln(\beta^*)]^2} + (n-1)*(\beta^*+T_*)^{-2}. \tag{6.10}$$

Indeed, the limiting value (6.10) *is supremal only, and for any $\beta > 0$ and $\boldsymbol{\tau} \in \mathfrak{T}_n$, we have $\mathscr{I}(T_*\mathbf{1}_n,\beta) = (\frac{\partial}{\partial\beta}\ln\mathcal{V}[X(T_*)])^2 \leq \mathscr{I}(\boldsymbol{\tau},\beta)$. Consequently, there is no locally optimal design for the estimation of β under the process governed by* (6.9)*.*

The previous example demonstrates that difficulties with the existence of locally optimal designs might emerge when the Fisher information $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ is not continuous at the boundary $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$. This fact underlines the importance of the asymptotic properties of the Fisher information matrix for conclusions on the existence of locally optimal designs.

By specifying the information function Φ, i.e., if the measure of information is defined by $\Phi[\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)]$, we consider the problem of the existence of a Φ-optimal design on $\mathfrak{T}_n$. The particular choice of the information function Φ may, however, suppress the effect of the discontinuity of the Fisher information matrix, and thus different information functions might lead to different conclusions. In this section we discuss the existence of optimal designs in a stricter way: we say that an optimal n-point sampling design "exists in the strong sense" if for any boundary design $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$, there exists a design $\boldsymbol{\tau} \in \mathfrak{T}_n$, such that for any sequence of designs $\{\boldsymbol{\tau}^{(k)}\}_k$ on $\mathfrak{T}_n$ with $\lim_{k\to\infty} \boldsymbol{\tau}^{(k)} = \boldsymbol{\tau}_0$, the Fisher information matrix $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ Loewner dominates the matrix $\lim_{k\to\infty} \mathscr{I}(\boldsymbol{\tau}^{(k)}, \boldsymbol{\vartheta}^*)$. Clearly, if the condition for the existence of optimal designs in the stronger sense is satisfied, then an optimal sampling design exists regardless of the choice of information function.

We can exploit the statement of Lemma 4 to get a detailed view of the limiting behaviour of the Fisher information matrix for the underlying stochastic differential equation (6.1):

Lemma 7. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by stochastic differential equation* (6.1) *and $\boldsymbol{\vartheta}^*$ be the prior guess at the true value of $\boldsymbol{\vartheta}$. Then, for any constants π_1 and π_2,*

$$\lim_{\Delta\to 0} \left(\mathcal{E}_{X(t+\pi_1\Delta)}\left[\mathscr{I}_{X(t+\pi_2\Delta)|X(t+\pi_1\Delta)}(\boldsymbol{\vartheta}^*)\right] - \mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)(\pi_2 - \pi_1)\Delta - \mathscr{O}(t, \boldsymbol{\vartheta}^*)\right) = \mathbf{0}_{m\times m},$$

where

$$\mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*) = \left.\frac{\frac{\partial f_{\boldsymbol{\vartheta}}(t,x)}{\partial \boldsymbol{\vartheta}}\Big|_{x=\mathcal{E}[X(t)]} \frac{\partial f_{\boldsymbol{\vartheta}}(t,x)}{\partial \boldsymbol{\vartheta}^\mathsf{T}}\Big|_{x=\mathcal{E}[X(t)]} + \frac{\partial b_{\boldsymbol{\vartheta}}(t)}{\partial \boldsymbol{\vartheta}} \frac{\partial b_{\boldsymbol{\vartheta}}(t)}{\partial \boldsymbol{\vartheta}^\mathsf{T}} \mathcal{V}_{\boldsymbol{\vartheta}}[X(t)]}{\sigma^2_\beta(t)}\right|_{\boldsymbol{\vartheta}^*}$$

and

$$\mathscr{O}(t, \boldsymbol{\vartheta}^*) = \frac{1}{2} \left.\frac{\partial \ln \sigma^2_\beta(t)}{\partial \boldsymbol{\vartheta}} \frac{\partial \ln \sigma^2_\beta(t)}{\partial \boldsymbol{\vartheta}^\mathsf{T}}\right|_{\boldsymbol{\vartheta}^*}.$$

Proof. Since the distribution of $X(t + \pi_2\Delta) \mid X(t + \pi_2\Delta)$ is Gaussian, we can write

$$\begin{aligned}\mathcal{E}_{X(t+\pi_1\Delta)}[\mathscr{I}_{X(t+\pi_2\Delta)|X(t+\pi_1\Delta)}(\boldsymbol{\vartheta}^*)] = \\ \frac{\mathcal{E}_{X(t+\pi_1\Delta)}\left[\frac{\partial \mathcal{E}[X(t+\pi_2\Delta)|X(t+\pi_1\Delta)]}{\partial \boldsymbol{\vartheta}} \frac{\partial \mathcal{E}[X(t+\pi_2\Delta)|X(t+\pi_1\Delta)]}{\partial \boldsymbol{\vartheta}^\mathsf{T}}\right]}{\mathcal{V}[X(t + \pi_2\Delta) \mid X(t + \pi_1\Delta)]} \\ + \frac{1}{2} \frac{\partial \ln \mathcal{V}[X(t + \pi_2\Delta) \mid X(t + \pi_1\Delta)]}{\partial \boldsymbol{\vartheta}} \frac{\partial \ln \mathcal{V}[X(t + \pi_2\Delta) \mid X(t + \pi_1\Delta)]}{\partial \boldsymbol{\vartheta}^\mathsf{T}}. \quad (6.11)\end{aligned}$$

Our goal is to find the Taylor expansions of $\mathcal{E}_{X(t+\pi_1\Delta)}[\mathscr{I}_{X(t+\pi_2\Delta)|X(t+\pi_1\Delta)}(\boldsymbol{\vartheta}^*)]$ at $\Delta = 0$. From the variance given in (6.5), we have

$$\mathcal{V}[X(t + \pi_2\Delta) \mid X(t + \pi_1\Delta)] = \sigma^2(t)(\pi_2 - \pi_1)\Delta + \mathrm{o}(\Delta), \quad (6.12)$$

where $\mathrm{o}(\Delta)/\Delta \to 0$ as $\Delta \to 0$. Since $\ln[\sigma^2(t)(\pi_2 - \pi_1)\Delta] = \ln \sigma^2(t) + \ln[(\pi_2 - \pi_1)\Delta]$, we get, for $\Delta \to 0$, that

$$\frac{\partial \ln \mathcal{V}[X(t + \pi_2\Delta) \mid X(t + \pi_1\Delta)]}{\partial \boldsymbol{\vartheta}} \to \frac{\partial \ln \sigma^2(t)}{\partial \boldsymbol{\vartheta}},$$

which yields the formula for $\mathscr{O}(t,\boldsymbol{\vartheta}^*)$. And now for the harder part. After some algebraic manipulations based on (6.4), we can write the conditioned expectation $\mathcal{E}[X(t+\pi_2\Delta) \mid X(t+\pi_1\Delta)]$ as follows:

$$\mathcal{E}[X(t+\pi_2\Delta) \mid X(t+\pi_1\Delta)] = \mathrm{e}^{B(t+\pi_2\Delta)-B(t)}\mathcal{E}[X(t)] + \int_t^{t+\pi_2\Delta} \mathrm{e}^{B(t+\pi_2\Delta)-B(\nu)}a(\nu)\mathrm{d}\nu + \mathrm{e}^{B(t+\pi_2\Delta)-B(t+\pi_1\Delta)}\left(X(t+\pi_1\Delta) - \mathcal{E}[X(t+\pi_1\Delta)]\right).$$

By the differentiation of $\mathcal{E}[X(t+\pi_2\Delta) \mid X(t+\pi_1\Delta)]$ with respect to $\boldsymbol{\vartheta}$, we obtain

$$\begin{aligned}\frac{\partial\mathcal{E}[X(t+\pi_2\Delta) \mid X(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}} &= \frac{\partial[B(t+\pi_2\Delta)-B(t)]}{\partial\boldsymbol{\vartheta}}\mathrm{e}^{B(t+\pi_2\Delta)-B(t)}\mathcal{E}[X(t)] \\ &\quad + \mathrm{e}^{B(t+\pi_2\Delta)-B(t)}\frac{\partial\mathcal{E}[X(t)]}{\partial\boldsymbol{\vartheta}} + \frac{\partial}{\partial\boldsymbol{\vartheta}}\int_t^{t+\pi_2\Delta}\mathrm{e}^{B(t+\pi_2\Delta)-B(\nu)}a(\nu)\mathrm{d}\nu \\ &\quad + \frac{\partial[B(t+\pi_2\Delta)-B(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}\mathrm{e}^{B(t+\pi_2\Delta)-B(t+\pi_1\Delta)}\left(X(t+\pi_1\Delta)-\mathcal{E}[X(t+\pi_1\Delta)]\right) \\ &\quad - \mathrm{e}^{B(t+\pi_2\Delta)-B(t+\pi_1\Delta)}\frac{\partial\mathcal{E}[X(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}. \quad (6.13)\end{aligned}$$

In the next step we express $\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathcal{E}[X(t+\pi_1\Delta)]$ in terms of $\mathcal{E}[X(t)]$ and $\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathcal{E}[X(t)]$. It follows from the formula for the mean value (6.4) that

$$\mathcal{E}[X(t)] = \mathrm{e}^{B(t)-B(t+\pi_1\Delta)}\mathcal{E}[X(t+\pi_1\Delta)] + \int_{t+\pi_1\Delta}^t \mathrm{e}^{B(t)-B(\nu)}a(\nu)\mathrm{d}\nu,$$

and, taking the derivative with respect to $\boldsymbol{\vartheta}$ and performing some calculations, yields

$$\begin{aligned}\frac{\partial\mathcal{E}[X(t)]}{\partial\boldsymbol{\vartheta}} &= \frac{\partial[B(t)-B(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}\mathcal{E}[X(t)] \\ &\quad - \frac{\partial[B(t)-B(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}\int_{t+\pi_1\Delta}^t \mathrm{e}^{B(t)-B(\nu)}a(\nu)\mathrm{d}\nu \\ &\quad + \mathrm{e}^{B(t)-B(t+\pi_1\Delta)}\frac{\partial\mathcal{E}[X(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}} + \frac{\partial}{\partial\boldsymbol{\vartheta}}\int_{t+\pi_1\Delta}^t \mathrm{e}^{B(t)-B(\nu)}a(\nu)\mathrm{d}\nu, \quad (6.14)\end{aligned}$$

where we used the relation

$$\mathrm{e}^{B(t)-B(t+\pi_1\Delta)}\mathcal{E}[X(t+\pi_1\Delta)] = \mathcal{E}[X(t)] - \int_{t+\pi_1\Delta}^t \mathrm{e}^{B(t)-B(\nu)}a(\nu)\mathrm{d}\nu.$$

Extracting $\mathrm{e}^{B(t)-B(t+\pi_1\Delta)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathcal{E}[X(t+\pi_1\Delta)]$ from the relation (6.14) and subsequently setting

it to expression (6.13) leads to

$$\begin{aligned}\frac{\partial\mathcal{E}[X(t+\pi_2\Delta)\mid X(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}} &= \frac{\partial[B(t+\pi_2\Delta)-B(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}\mathrm{e}^{B(t+\pi_2\Delta)-B(t)}\mathcal{E}[X(t)]\\ &+\frac{\partial}{\partial\boldsymbol{\vartheta}}\int_t^{t+\pi_2\Delta}\mathrm{e}^{B(t+\pi_2\Delta)-B(\nu)}a(\nu)\mathrm{d}\nu+\mathrm{e}^{B(t+\pi_2\Delta)-B(t)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\int_{t+\pi_1\Delta}^{t}\mathrm{e}^{B(t)-B(\nu)}a(\nu)\mathrm{d}\nu\\ &+\frac{\partial[B(t+\pi_2\Delta)-B(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}\mathrm{e}^{B(t+\pi_2\Delta)-B(t+\pi_1\Delta)}\left(X(t+\pi_1\Delta)-\mathcal{E}[X(t+\pi_1\Delta)]\right)\\ &-\frac{\partial[B(t)-B(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}\int_{t+\pi_1\Delta}^{t}\mathrm{e}^{B(t+\pi_2\Delta)-B(\nu)}a(\nu)\mathrm{d}\nu\\ &=\frac{\partial a(t)}{\partial\boldsymbol{\vartheta}}(\pi_2-\pi_1)\Delta+\frac{\partial b(t)}{\partial\boldsymbol{\vartheta}}\mathcal{E}[X(t)](\pi_2-\pi_1)\Delta\\ &+\frac{\partial b(t)}{\partial\boldsymbol{\vartheta}}\left(X(t+\pi_1\Delta)-\mathcal{E}[X(t+\pi_1\Delta)]\right)(\pi_2-\pi_1)\Delta+\mathrm{o}(\Delta)\\ &=\left.\frac{\partial f(t,x)}{\partial\boldsymbol{\vartheta}}\right|_{x=\mathcal{E}[X(t)]}+\frac{\partial b(t)}{\partial\boldsymbol{\vartheta}}\left(X(t+\pi_1\Delta)-\mathcal{E}[X(t+\pi_1\Delta)]\right)(\pi_2-\pi_1)\Delta+\mathrm{o}(\Delta).\end{aligned}$$

The second equality results from the Taylor expansion of $B(t)$'s and integrals at $\Delta = 0$. Evidently,

$$\begin{aligned}&\mathcal{E}_{X(t+\pi_1\Delta)}\left[\frac{\partial\mathcal{E}[X(t+\pi_2\Delta)\mid X(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}}\frac{\partial\mathcal{E}[X(t+\pi_2\Delta)\mid X(t+\pi_1\Delta)]}{\partial\boldsymbol{\vartheta}^\mathsf{T}}\right]=\\ &\left.\frac{\partial f(t,x)}{\partial\boldsymbol{\vartheta}}\right|_{x=\mathcal{E}[X(t)]}\left.\frac{\partial f(t,x)}{\partial\boldsymbol{\vartheta}^\mathsf{T}}\right|_{x=\mathcal{E}[X(t)]}(\pi_2-\pi_1)^2\Delta^2+\frac{\partial b(t)}{\partial\boldsymbol{\vartheta}}\frac{\partial b(t)}{\partial\boldsymbol{\vartheta}^\mathsf{T}}\mathcal{V}[X(t)](\pi_2-\pi_1)^2\Delta^2.\end{aligned}\tag{6.15}$$

since $\mathcal{V}[X(t+\pi_1\Delta)] = \mathcal{V}[X(t)] + \mathrm{o}(1)$. By setting (6.15) and (6.12) to (6.11), we obtain the formula for $\mathscr{I}_\infty(t,\boldsymbol{\vartheta}^*)$. □

Since the volatility function $\sigma(t)$ depends solely on the parameter β, due to Lemma 7 the only element of the Fisher information matrix potentially not continuous at the boundary designs is $\{\mathscr{I}(\boldsymbol{\tau},\boldsymbol{\vartheta}^*)\}_{\beta\beta}$. Consequently, we obtain the following result:

Theorem 5. *If β is not an unknown parameter of stochastic differential equation* (6.1)*, then an optimal sampling design exists in the strong sense for the estimation of $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_1$ (for any n and $\boldsymbol{\vartheta}^*$).*

Proof. It follows from Lemma 6 that any neighbourhood of a boundary design $\boldsymbol{\tau}_0$ contains a feasible design $\boldsymbol{\tau}$ with $\mathscr{I}(\boldsymbol{\tau},\boldsymbol{\vartheta}^*) \succeq_\mathrm{L} \mathscr{I}(\boldsymbol{\tau}_0,\boldsymbol{\vartheta}^*)$. If β is not an unknown parameter, then Lemma 7 yields a continuity of the Fisher information matrix at the boundary, i.e., $\lim_{\boldsymbol{\tau}\to\boldsymbol{\tau}_0}\mathscr{I}(\boldsymbol{\tau},\boldsymbol{\vartheta}^*) = \mathscr{I}(\boldsymbol{\tau}_0,\boldsymbol{\vartheta}^*)$. □

A technique used in the proof of Theorem 5 can also be applied when $\boldsymbol{\theta} = (\boldsymbol{\theta}_\mathcal{E}^\mathsf{T},\boldsymbol{\theta}_\mathcal{V}^\mathsf{T})^\mathsf{T}$ with $\frac{\partial}{\partial\boldsymbol{\theta}_\mathcal{V}}\mathcal{E}[X(t)] = \frac{\partial}{\partial\beta}\mathcal{E}[X(t)] = 0$ and $\frac{\partial}{\partial\boldsymbol{\theta}_\mathcal{E}}\mathcal{V}[X(t)] = \mathbf{0}$. Then the Fisher information about $(X_0,\boldsymbol{\theta}_\mathcal{E}^\mathsf{T})^\mathsf{T}$ arising from the Schur complement (see formula (6.20)) is equal to the block of the Fisher information matrix $\mathscr{I}(\boldsymbol{\tau},\boldsymbol{\vartheta}^*)$, which corresponds to $(X_0,\boldsymbol{\theta}_\mathcal{E}^\mathsf{T})^\mathsf{T}$. This block is, however, continuous, and, henceforth, an optimal design for the estimation of $(X_0,\boldsymbol{\theta}_\mathcal{E}^\mathsf{T})^\mathsf{T}$ exists in the strong sense.

In a more general situation where the correlation between the observations is present, the question of the existence of optimal sampling designs is a complex problem. The main

reason is that the Fisher information matrix and the subsequent optimisation problems are typically non-convex in the design points.

One possibility for verifying whether the optimal value is achieved by a convergence to a boundary design is to focus on the local behaviour of the Fisher information matrix at the boundary points. For this purpose we can use a modification of the traditional concept of a directional derivative, which has turned out to be a useful tool in the theory of optimal experimental designs with uncorrelated observations; see, e.g., [77].

The present setup, however, puts limitations on the choice of feasible directions. For any boundary point $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$ we can construct a feasible direction $\boldsymbol{\tau} = (t_1, \ldots, t_n)^\mathsf{T}$ by taking $t_i = \{\boldsymbol{\tau}_0\}_i + \pi_i$, $i = 1, \ldots, n$, where $\pi_i = 0$ if $\{\boldsymbol{\tau}_0\}_i > \{\boldsymbol{\tau}_0\}_{i-1}$, and taking $\pi_i > \pi_{i-1} \geq 0$ if $\{\boldsymbol{\tau}_0\}_i = \{\boldsymbol{\tau}_0\}_{i-1}$ and $t_i > t_{i-1}$. We further require $\{\boldsymbol{\tau}_0\}_i + \pi_i > \{\boldsymbol{\tau}_0\}_{i-1} + \pi_{i-1}$. For the Fisher information matrix we can then write

$$\partial\mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\pi}, \boldsymbol{\vartheta}^*) = \lim_{\Delta \searrow 0} \frac{\mathscr{I}(\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}, \boldsymbol{\vartheta}^*) - \lim_{\Delta \searrow 0} \mathscr{I}(\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}, \boldsymbol{\vartheta}^*)}{\Delta}. \tag{6.16}$$

The vector $\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}$ has strictly increasing components and converges to $\boldsymbol{\tau}_0$ as $\Delta \to 0$. In addition, the directional derivative $\partial\mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\pi}, \boldsymbol{\vartheta}^*)$ is positive homogeneous in $\boldsymbol{\pi}$ (see formula (6.18) below), i.e., the condition $\{\boldsymbol{\tau}_0\}_i + \pi_i > \{\boldsymbol{\tau}_0\}_{i-1} + \pi_{i-1}$ does not have to be satisfied, since for any sufficiently small Δ, the vector $\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}$ is a vector of strictly increasing sampling times. We can, therefore, impose a weaker requirement on the direction $\boldsymbol{\pi}$:

$$\boldsymbol{\pi} = \begin{cases} 0, & \{\boldsymbol{\tau}_0\}_i > \{\boldsymbol{\tau}_0\}_{i-1} \\ \pi_i > \pi_{i-1}, & \{\boldsymbol{\tau}_0\}_i = \{\boldsymbol{\tau}_0\}_{i-1} \end{cases}. \tag{6.17}$$

The use of the directional derivative is obvious: if for any boundary point $\boldsymbol{\tau}_0$, there exists a direction $\boldsymbol{\pi}$ in line with (6.17) such that the matrix $\partial\mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\pi}, \boldsymbol{\vartheta}^*)$ is positive definite, then the supremum information is not reached when we converge to the boundary $\overline{\mathfrak{T}}_n$, regardless of the choice of information function, or in other words, an optimal design exists in the strong sense.

We can analogously take $\Phi[\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)]$ instead of $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ in the directional derivative (6.16) if the information criterion is specified.

Let us assume for a while that the initial value X_0 is a known quantity and that $\{\boldsymbol{\tau}_0\}_n < T^*$. From Lemmata 4 and 7 we obtain that for $\boldsymbol{\pi}$ defined by (6.17), we have

$$\lim_{\Delta \searrow 0} \mathscr{I}(\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}, \boldsymbol{\vartheta}^*) = \sum_{i=1, \pi_i = 0}^{n} \mathcal{E}_{X(\{\boldsymbol{\tau}_0\}_{i-1})} \left[\mathscr{I}_{X(\{\boldsymbol{\tau}_0\}_i)|X(\{\boldsymbol{\tau}_0\}_{i-1})}(\boldsymbol{\vartheta}^*) \right] + \sum_{i=1, \pi_i \neq 0}^{n} \mathscr{O}(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*),$$

where $X(t_0) = X(0)$. Next, if $\pi_i \neq 0$, then $\{\boldsymbol{\tau}_0\}_i = \{\boldsymbol{\tau}_0\}_{i-1}$ and, for any Δ sufficiently small, we get $\mathcal{E}_{X(\{\boldsymbol{\tau}_0\}_{i-1} + \pi_{i-1}\Delta)}[\mathscr{I}_{X(\{\boldsymbol{\tau}_0\}_i + \pi_i\Delta)|X(\{\boldsymbol{\tau}_0\}_{i-1} + \pi_{i-1}\Delta)}(\boldsymbol{\vartheta}^*)] = \mathscr{I}_\infty(t_i, \boldsymbol{\vartheta}^*)(\pi_i - \pi_{i-1})\Delta + \mathscr{O}(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*) + \mathrm{o}(\Delta)$. We can thus write

$$\mathscr{I}(\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}, \boldsymbol{\vartheta}^*) = \sum_{i=1, \pi_i = 0}^{n} \mathcal{E}_{X(\{\boldsymbol{\tau}_0\}_{i-1})} \left[\mathscr{I}_{X(\{\boldsymbol{\tau}_0\}_i)|X(\{\boldsymbol{\tau}_0\}_{i-1})}(\boldsymbol{\vartheta}^*) \right] + \sum_{i=1, \pi_i \neq 0}^{n} \mathscr{O}(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*) + \sum_{i=1, \pi_i \neq 0}^{n} \mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)(\pi_i - \pi_{i-1})\Delta + \mathrm{o}(\Delta),$$

where $\mathrm{o}(\Delta)/\Delta \to 0$ as $\Delta \to 0$. Consequently, for the directional derivative (6.16) we obtain

$$\partial\mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\pi}, \boldsymbol{\vartheta}^*) = \sum_{i=1,\pi_i\neq 0}^{n} \mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)(\pi_i - \pi_{i-1}). \tag{6.18}$$

Now we address the assumptions of the previous paragraph. Firstly, the assumption $\{\boldsymbol{\tau}_0\}_n < T^*$ can be treated by the fact that

$$\lim_{\Delta\to 0} \mathcal{E}_{X(\{\boldsymbol{\tau}_0\}_i+\pi_{i-1}\Delta)}[\mathscr{I}_{X(\{\boldsymbol{\tau}_0\}_i+\pi_i\Delta)|X(\{\boldsymbol{\tau}_0\}_i+\pi_{i-1}\Delta)}(\boldsymbol{\vartheta}^*)] = \lim_{\Delta\to 0} \mathcal{E}_{X(\{\boldsymbol{\tau}_0\}_i-\pi_i\Delta)}[\mathscr{I}_{X(\{\boldsymbol{\tau}_0\}_i-\pi_{i-1}\Delta)|X(\{\boldsymbol{\tau}_0\}_i-\pi_i\Delta)}(\boldsymbol{\vartheta}^*)], \tag{6.19}$$

that is, we perform an infinitesimal perturbation in the "opposite direction". Secondly, if X_0 is an unknown parameter, then the first row and the first column of $\mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)$ are zero, and we therefore focus on the submatrix of $\mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)$ corresponding to parameters $\boldsymbol{\theta}$ and β. Finally, since the directional derivative is positive homogeneous in $\boldsymbol{\pi}$, its rank does not depend on $\boldsymbol{\pi}$, and we can substitute the value of $\pi_i - \pi_{i-1}$ for any positive constant.

Theorem 6. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by stochastic differential equation* (6.1) *and $\boldsymbol{\vartheta}^*$ be the prior guess at the true value of $\boldsymbol{\vartheta}$. If for a boundary design $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$, there exist a vector $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\boldsymbol{\tau}_0)$ of non-negative constants such that the matrix*

$$\mathcal{Z}(\boldsymbol{\tau}_0, \boldsymbol{\alpha}, \boldsymbol{\vartheta}^*) = \sum_{i=1,\{\boldsymbol{\tau}_0\}_i=\{\boldsymbol{\tau}_0\}_{i-1}}^{n} \alpha_i \begin{pmatrix} \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\theta}\beta} \\ \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\beta\boldsymbol{\theta}} & \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\beta\beta} \end{pmatrix}$$

is positive definite, then the supremum information is not attained when we converge to $\boldsymbol{\tau}_0$, regardless of the choice of information function.

Note that the matrix $\mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)$ is of a rank of at most two. The subadditivity of the matrix rank (see, e.g., inequality 0.4.5(d) by Horn and Johnson [44]) implies that the rank of $\mathcal{Z}(\boldsymbol{\tau}_0, \boldsymbol{\alpha}, \boldsymbol{\vartheta}^*)$ is at most two times the cardinality of the set $\{i : \{\boldsymbol{\tau}_0\}_i \neq \{\boldsymbol{\tau}_0\}_{i-1}\}$. Theorem 6 is henceforth useful in the verification of the existence of optimal sampling designs in the strong sense when $\dim(\boldsymbol{\theta}) = 1$ (that is, $\dim(\boldsymbol{\vartheta}) = 3$), where a convergence to a one-point design can be checked; nevertheless, it might provide some guidance for higher dimensions of $\boldsymbol{\theta}$.

We should conclude this section by giving a few points concerning the statistical properties of estimators. The limiting behaviour of the conditioned Fisher information matrix $\mathcal{E}_{X(t-\Delta)}\left[\mathscr{I}_{X(t)|X(t-\Delta)}(\boldsymbol{\vartheta}^*)\right]$ given in Lemma 7 indicates that for $\Delta \to 0$, the only nonzero element of $\mathcal{E}_{X(t-\Delta)}[\mathscr{I}_{X(t)|X(t-\Delta)}(\boldsymbol{\vartheta}^*)]$ corresponds to parameter β. Consequently, with the increasing number of observations at distinct sampling times from the experimental domain $\mathcal{D}$, the amount of information about β increases above any limits, and so β can be estimated consistently; see the paper by Crowder [15] for a detailed discussion of the consistency of maximum likelihood estimators. This phenomenon has a natural explanation: Unlike X_0 or $\boldsymbol{\theta}$, the parameter β in equation (6.1) is connected with the differential of the Wiener process $\{W(t)\}_{t\geq 0}$, which has a nonzero quadratic variation and is characterized by the fractal property known as Brownian scaling. An analogous formulation of this result can be found in the literature on stochastic differential equations, see, e.g., the monograph of Iacus [45] for a brief survey, but it is also noted in selected papers on inference in regression problems with correlated observations; see, for instance, a Pázman's early paper [76].

6.3 Ultimate efficiency of designs

The formulation of the ultimate efficiency proposed by Pázman [79] is applicable only if $\lim_{n\to\infty} \Phi[\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)]$ is finite. In our situation, where the parameter β is consistently estimable, if $\{\boldsymbol{\tau}^{(n)}\}_{n\geq m} \in \mathfrak{C}_{\mathcal{D}}$, then, for $n \to \infty$, the Fisher information matrix $\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)$ converges to a matrix $\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)$ with only $m-1$ eigenvalues bounded, and thus the measure of information might tend to infinity as well.

Henceforth, we focus on the subparameter $\boldsymbol{\vartheta}_{\mathrm{I}}$: in general, if $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_{\mathrm{I}}^{\mathsf{T}}, \boldsymbol{\vartheta}_{\mathrm{II}}^{\mathsf{T}})^{\mathsf{T}}$ is a partition of the unknown parameter and $\boldsymbol{\tau} \in \mathfrak{T}_n$, the Fisher information matrix $\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ corresponding to $\boldsymbol{\vartheta}_{\mathrm{I}}$ is given by the Schur complement

$$\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) = \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}} - \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}^{-}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{I}}}, \quad (6.20)$$

where $\{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}^{-}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}}$ is an arbitrary pseudo-inverse of $\{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}}$. In addition, on account of Theorem 3.1(b) by Li and Mathias [63], the Schur complements preserve Loewner ordering; hence, we have the following result:

Proposition 50. *If an optimal sampling design for the estimation of $\boldsymbol{\vartheta}$ exists in the strong sense, then an optimal sampling design for the estimation of $\boldsymbol{\vartheta}_{\mathrm{I}}$ exists in the strong sense.*

In the sequel we extend the definition of the ultimate efficiency suggested by Pázman [79].

Definition 27. *Let $\{\boldsymbol{\tau}^{(n)}\}_{n\geq m} \in \mathfrak{C}_{\mathcal{D}}$ and let $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_{\mathrm{I}}^{\mathsf{T}}, \boldsymbol{\vartheta}_{\mathrm{II}}^{\mathsf{T}})^{\mathsf{T}}$ be a partition of the unknown parameter, for which*

$$\lambda_{\max}\left(\{\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}\right) < \infty,$$

where $\lambda_{\max}$ denotes the largest eigenvalue, and

$$\{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}^{-}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{I}}} \to \mathbf{0}_{\dim(\boldsymbol{\vartheta}_{\mathrm{I}})\times\dim(\boldsymbol{\vartheta}_{\mathrm{I}})}, \quad (6.21)$$

as $n \to \infty$. Then, the (local) ultimate efficiency of a design $\boldsymbol{\tau} \in \mathfrak{T}_n$ for the estimation of $\boldsymbol{\vartheta}_{\mathrm{I}}$ with respect to an information function Φ, briefly, the (local) ultimate Φ-efficiency, is the ratio

$$\mathrm{ueff}(\boldsymbol{\tau} \mid \Phi, \boldsymbol{\vartheta}^*) = \frac{\Phi[\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)]}{\Phi[\{\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}]}.$$

We should provide a little more of the discussion on the condition (6.21) in Definition 27. For any sequence of designs $\{\boldsymbol{\tau}^{(n)}\}_n \in \mathfrak{C}_{\mathcal{D}}$ covering $\mathcal{D}$, the Schur complement (6.20) implies $\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*) \preceq_{\mathrm{L}} \{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}} \to \{\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}$ and, as a consequence of Lemma 6, $\Phi[\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)] \leq \Phi[\{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}] \nearrow \Phi[\{\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)]$. On the other hand, the condition (6.21) ensures that $\Phi[\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)] \to \Phi[\{\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}]$ as $n \to \infty$. Hence, $\mathrm{ueff}(\boldsymbol{\tau} \mid \Phi, \boldsymbol{\vartheta}^*) \leq 1$ and $\mathrm{ueff}(\boldsymbol{\tau}^{(n)} \mid \Phi, \boldsymbol{\vartheta}^*) \to 1$ as $n \to \infty$. The concept of the ultimate efficiency is also properly defined for instances where the optimal designs do not exist, because the condition (6.21) remains valid and $\{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}$ is continuous on $\overline{\mathfrak{T}}_n$.

From Lemmata 4 and 7 we obtain one of the main results:

Theorem 7. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by stochastic differential equation* (6.1)*, and let $\{\boldsymbol{\tau}^{(n)}\}_{n\geq m} \in \mathfrak{C}_{\mathcal{D}}$. Then*

$$\lim_{n\to\infty} \left(\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*) - \mathscr{I}_\infty(\boldsymbol{\vartheta}^*) - \sum_{i=2}^{n} \mathscr{O}(t_i, \boldsymbol{\vartheta}^*) \right) = \mathbf{0}_{m\times m},$$

where

$$\mathscr{I}_\infty(\boldsymbol{\vartheta}^*) = \left.\frac{\frac{\partial \mathcal{E}[X(T_*)]}{\partial \boldsymbol{\vartheta}} \frac{\partial \mathcal{E}[X(T_*)]}{\partial \boldsymbol{\vartheta}^\mathsf{T}}}{\mathcal{V}[X(T_*)]}\right|_{\boldsymbol{\vartheta}^*} + \frac{1}{2}\left.\frac{\partial \ln \mathcal{V}[X(T_*)]}{\partial \boldsymbol{\vartheta}} \frac{\partial \ln \mathcal{V}[X(T_*)]}{\partial \boldsymbol{\vartheta}^\mathsf{T}}\right|_{\boldsymbol{\vartheta}^*} + \int_{T_*}^{T^*} \mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)\mathrm{d}t, \tag{6.22}$$

and $\mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)$ and $\mathscr{O}(t, \boldsymbol{\vartheta}^*)$ are defined in Lemma 7.

Since the parameter β is consistently estimable, using the previous theorem we can, in line with Definition 27, compute the ultimate Φ-efficiency of designs for parameter subvector $\boldsymbol{\vartheta}_\mathrm{I} = (X_0, \boldsymbol{\theta}^\mathsf{T})^\mathsf{T}$.

6.4 Ornstein-Uhlenbeck type processes

In the previous sections we presented an analysis of a generalised form of the Ornstein-Uhlenbeck process described by stochastic differential equation (6.1), which enables us to evaluate the quality of the sampling designs. A reasonable question is whether we can also use the results for processes other than those governed by equation (6.1).

The motivation comes from the Fisher-Neymann factorization theorem; see, for instance, Theorem 6.5 in [61]. More precisely, if we apply a sufficient statistic to the measurements, then the Fisher information matrix remains unchanged. Henceforth, we can define the following class of stochastic differential equations.

Definition 28. *Let $\mu_{\boldsymbol{\theta}\beta}(t, y)$ and $\gamma_\beta(t, y)$ be sufficiently smooth, and let a process $\{Y(t)\}_{t\geq 0}$ be governed by the stochastic differential equation*

$$\mathrm{d}Y(t) = \mu_{\boldsymbol{\theta},\beta}(t, Y(t))\mathrm{d}t + \gamma_\beta(t, Y(t))\mathrm{d}W(t).$$

If there exist sufficiently smooth functions $a_{\boldsymbol{\theta},\beta}(t)$, $b_{\boldsymbol{\theta},\beta}(t)$, $\sigma_\beta(t)$ and $\varphi(t, y)$, where φ is bijection in y and $\frac{\partial \varphi(t,y)}{\partial \boldsymbol{\vartheta}} = \mathbf{0}_m$ for all $t \geq 0$ and $y \in \mathbb{R}$, such that the process $\{X(t)\}_{t\geq 0} = \{\varphi(t, Y(t))\}_{t\geq 0}$ is governed by the equation

$$\mathrm{d}X(t) = [a_{\boldsymbol{\theta},\beta}(t) + b_{\boldsymbol{\theta},\beta}(t)X(t)]\mathrm{d}t + \sigma_\beta(t)\mathrm{d}W(t),$$

then we say that the process $\{Y(t)\}_{\geq 0}$ is an Ornstein-Uhlenbeck type process with associated coefficients $a(t)$, $b(t)$ and $\sigma(t)$. We denote this fact by $\{Y(t)\} \in \mathcal{OU}_{a(t),b(t),\sigma(t)}$.

A candidate for the desired sufficient statistic, which transforms the volatility of the original process to the volatility of the desired form, is

$$\varphi(t, y) = \int \frac{\sigma(t)}{\gamma(t, y)}\mathrm{d}y.$$

The condition that $\frac{\partial \varphi(t,y)}{\partial \boldsymbol{\vartheta}} = \mathbf{0}_m$ for all $t \geq 0$ and $y \in \mathbb{R}$ might not be easy to verify in advance, because we do not know the form of $\sigma(t)$. Nonetheless, if we can write the diffusion term in the separable form $\gamma(t, y) = \sigma(t)g(t, y)$, where $\frac{\partial g(t,y)}{\partial \boldsymbol{\vartheta}} = \mathbf{0}_m$ for all $t \geq 0$ and $y \in \mathbb{R}$, then φ is a sufficient statistic. We remark that φ depends on a reciprocal of $\gamma(t, y)$; thus we might need to impose further positivity conditions on the domain interior of $Y(t)$ for all t, which is outside the scope of the presented paper.

In the sequel we propose a way to verify whether a given process is of the Ornstein-Uhlenbeck type. Let $\psi(t, y)$ be the inverse function of $\varphi(t, y)$, that is, $\psi(t, \varphi(t, y)) = y$. Then $\{Y(t)\}_{t\geq 0} = \{\psi(t, X(t))\}_{t\geq 0}$. Itō's lemma implies

$$\left.\frac{\partial \psi}{\partial t}\right|_{x=\varphi(t,y)} + \left.\frac{\partial \psi}{\partial x}\right|_{x=\varphi(t,y)} f(t, \varphi(t, y)) + \frac{1}{2}\left.\frac{\partial^2 \psi}{\partial x^2}\right|_{x=\varphi(t,y)} \sigma^2(t) = \mu(t, y). \tag{6.23}$$

By setting the relations for inverse functions

$$\left.\frac{\partial\psi}{\partial t}\right|_{x=\varphi(t,y)} = -\frac{\partial\varphi}{\partial t}\Big/\frac{\partial\varphi}{\partial y},\ \left.\frac{\partial\psi}{\partial x}\right|_{x=\varphi(t,y)} = 1\Big/\frac{\partial\varphi}{\partial y},\ \left.\frac{\partial^2\psi}{\partial x^2}\right|_{x=\varphi(t,y)} = -\frac{\partial^2\varphi}{\partial y^2}\Big/\left(\frac{\partial\varphi}{\partial y}\right)^3.$$

to equation (6.23), we obtain the following theorem:

Theorem 8. *Let the process $\{Y(t)\}_{t\geq 0}$ be driven by the stochastic differential equation*

$$\mathrm{d}Y(t) = \mu(t, Y(t))\mathrm{d}t + \sigma(t)g(t, Y(t))\mathrm{d}W(t),$$

where the functions $\mu(t,y)$ and $g(t,y)$ are sufficiently smooth, and $\frac{\partial g(t,y)}{\partial\boldsymbol{\theta}} = 0$ for all $t \geq 0$ and $y \in \mathbb{R}$. If the condition

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\frac{\mathrm{d}y}{g(t,y)} = a(t) + b(t)\int\frac{\mathrm{d}y}{g(t,y)} + \frac{1}{2}\sigma^2(t)\frac{\partial g(t,y)}{\partial y} - \frac{\mu(t,y)}{g(t,y)} \tag{6.24}$$

is satisfied for some functions $a_{\boldsymbol{\theta},\beta}(t)$ and $b_{\boldsymbol{\theta},\beta}(t)$, then $\{Y(t)\}_{t\geq 0} \in \mathcal{OU}_{a_{\boldsymbol{\theta},\beta}(t),b_{\boldsymbol{\theta},\beta}(t),\sigma_\beta(t)}$.

In some instances we can consider a process governed by an autonomous stochastic differential equation of the form

$$\mathrm{d}Y(t) = \mu_{\boldsymbol{\theta},\beta}(Y(t))\mathrm{d}t + \sigma_\beta g(Y(t))\mathrm{d}W(t), \tag{6.25}$$

where the structure of the drift function $\mu(y)$ is usually based on an essential theory in the given research field. On the contrary, the choice of $g(y)$ might be artificial; we can choose a diffusion that fits some arrangements. By a differentiation of equation (6.24) with respect to y, we get

Corollary 1. *Let the process $\{Y(t)\}_{t\geq 0} \in \mathcal{OU}_{a_{\boldsymbol{\theta},\beta},b_{\boldsymbol{\theta},\beta},\sigma_\beta}$ be driven by the stochastic differential equation* (6.25) *and let $\mu(y)$ be given. Then $g(y)$ solves the ordinary differential equation*

$$\left(b - \frac{\partial\mu(y)}{\partial y}\right)g(y) + \mu(y)\frac{\partial g(y)}{\partial y} + \frac{1}{2}\sigma^2 g^2(y)\frac{\partial^2 g(y)}{\partial y^2} = 0.$$

If the solution $g(y)$ at least approximately corresponds to the experimental setting, we can use the proposed methodology for an assessment of the design's quality. A practical demonstration is presented in the next section.

6.5 Example: Gompertz model of tumour growth

Tumour growth models play an important role in therapeutic guidance. Gompertz [36] proposed in his pioneering paper a growth model, which became the basis for many studies in cancer research. Various modifications of this growth law have been introduced; we refer the reader to Norton [72] and Speer et al. [100] for a brief survey. Although the work of Cameron [10] demonstrates that the Gompertz model is not the best choice for breast cancer, it seems to fit the data for multiple myeloma well; see Sullivan and Salmon [102].

In the Gompertz growth model the expected size of the tumour $Y_{\theta_1,\theta_2}(t)$ is the solution to the ordinary differential equation

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = \theta_2 Y(t) - \theta_1 Y(t)\ln Y(t),\ Y(0) = Y_0 \text{ known}, \tag{6.26}$$

where the growth deceleration factor θ_1 and the intrinsic growth rate θ_2 are unknown parameters of interest, and the initial size Y_0 is usually obtained from a first-time detection. The solution to equation (6.26) is known in the explicit closed form

$$Y_{\theta_1,\theta_2}(t) = \mathrm{e}^{\mathrm{e}^{-\theta_1 t}\ln Y_0 + (1-\mathrm{e}^{-\theta_1 t})\frac{\theta_2}{\theta_1}}, \tag{6.27}$$

which might be a factor explaining its popularity in applications.

To improve the quality of the statistical inference, it is important to choose a proper experimental design. Locally D-optimal designs for model (6.26) were studied by Li [64], who assumed that the observations $\tilde{Y}_{ij}$ satisfy the regression model

$$\tilde{Y}_{ij} = Y_{\theta_1,\theta_2}(t_i) + \varepsilon_{ij},\ i = 1,\dots,n,\ j = 1,\dots,k_i.$$

Here, $\tilde{Y}_{ij}$ are the observations of the tumour size; t_i's are the design times at which we perform k_i replications; and ε_{ij}'s are independent homoscedastic errors.

In this section we consider a variant of the Gompertz growth model, where the size of the tumour $\{Y(t)\}_{t\geq 0}$ is described by the stochastic differential equation

$$\mathrm{d}Y(t) = [\theta_2 Y(t) - \theta_1 Y(t)\ln Y(t)]\mathrm{d}t + \beta Y(t)\mathrm{d}W(t),\ Y(0) = Y_0 \text{ known}, \tag{6.28}$$

see Lo [66] for more detail. Remark that the structure of the volatility in equation (6.28) implies small fluctuations in the tumour size for small tumours, but, to the contrary, if the size of the tumour is large, then the fluctuations in the size of the tumour are large, too.

Compared to the model of Li [64], for the process $\{Y(t)\}_{t\geq 0}$ governed by equation (6.28), it has no point to replicate the observations, and so the amount of information about the parameters might be limited.

If we define $\{X(t)\}_{t\geq 0} = \{\ln Y(t)\}_{t\geq 0}$, then $\{X(t)\}_{t\geq 0}$ solves

$$\mathrm{d}X(t) = \left(\theta_2 - \frac{1}{2}\beta^2 - \theta_1 X(t)\right)\mathrm{d}t + \beta\mathrm{d}W(t).$$

That is, the process $\{Y(t)\}_{t\geq 0}$ is an Ornstein-Uhlenbeck type process with the associated coefficients

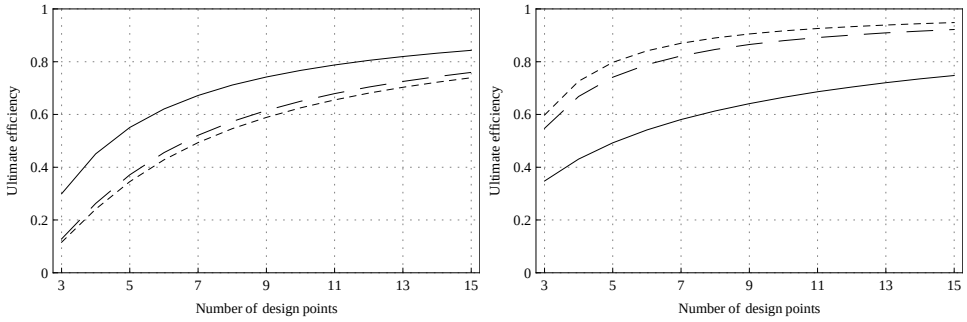
$$a_{\theta_2,\beta}(t) = \theta_2 - \frac{1}{2}\beta^2,\ b_{\theta_1}(t) = -\theta_1 \text{ and } \sigma_\beta(t) = \beta.$$

By setting $\boldsymbol{\theta} = (\theta_1,\theta_2)^\mathsf{T} = \boldsymbol{\vartheta}_\mathrm{I}$, we get the asymptotic Fisher information matrix for $\boldsymbol{\theta}$

$$\{\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\frac{\partial\mathcal{E}[X(T_*)]}{\partial\boldsymbol{\theta}}\frac{\partial\mathcal{E}[X(T_*)]}{\partial\boldsymbol{\theta}}}{\mathcal{V}[X(T_*)]} + \frac{\partial\ln\mathcal{V}[X(T_*)]}{\partial\boldsymbol{\theta}}\frac{\partial\ln\mathcal{V}[X(T_*)]}{\partial\boldsymbol{\theta}} + \frac{1}{\beta^2}\int_{T_*}^{T^*}\begin{pmatrix}\mathcal{E}^2[X(t)]+\mathcal{V}[X(t)] & -\mathcal{E}[X(t)]\\ -\mathcal{E}[X(t)] & 1\end{pmatrix}\mathrm{d}t,$$

where $\mathcal{E}[X(t)] = \mathrm{e}^{-\theta_1 t}\ln Y_0 + \frac{1}{\theta_1}(\theta_2 - \frac{1}{2}\beta^2)(1-\mathrm{e}^{-\theta_1 t})$, $\mathcal{V}[X(t)] = \frac{\beta^2}{2\theta_1}(1-\mathrm{e}^{-2\theta_1 t})$. We note that with the exception of the information contained in the first observation (which is zero if $T_* \to 0$), the asymptotic information depends only on the mean value and the variance of the process.

Figure 6.1 depicts the dependence of the ultimate efficiency of an equidistant n-point design on the number of design points. In this example we consider $\mathcal{D} = [1,2]$ and designs with $t_1 = 1$ and $t_n = 2$, and two different prior values of θ_1, θ_2 and β. By comparing the illustrative results (cf. E-optimality), it is evident that the nonlinearity of the regression



a) $\theta_1^* = 3, \theta_2^* = 1, \beta^* = 1$ b) $\theta_1^* = 1, \theta_2^* = 1, \beta^* = 3$

Figure 6.1. Ultimate efficiencies of equidistant n-point designs in experimental domain $\mathcal{D} = [1, 2]$ with $t_1 = 1$, $t_n = 2$, $Y_0 = 1$ and different prior estimates θ_1^*, θ_2^* and β^*. We consider the D-optimality criterion (——), E-optimality criterion (- - -) and A-optimality criterion (— —).

model makes the ultimate efficiency sensitive to the choice of the information function and prior guess at the true value of the unknown parameters.

The prior estimates of the true values of parameters θ_1 and θ_2 required for local linearization could be obtained from the previous experimental data by estimating, for instance, multiresponse regression models, which capture the relations between the estimates and different physiological factors.

Chapter 7

On the information contained in the trajectory of a process governed by a stochastic differential equation

Consider a general situation, where we observe an Itō process $\{X(t)\}_{t\geq 0}$ described by the stochastic differential equation

$$\mathrm{d}X(t) = f_{\boldsymbol{\theta},\beta}(t, X(t))\mathrm{d}t + \sigma_\beta(t, X(t))\mathrm{d}W(t). \tag{7.1}$$

We shall assume that there exists at least a weak solution to (7.1), its transition density kernel $p(x, s \mid y, t) = \frac{\partial}{\partial x}\Pr[X(t+s) < x \mid X(t) = y]$ and its first and second derivatives with respect to the unknown vector parameter $\boldsymbol{\vartheta} = (\boldsymbol{\theta}^\mathsf{T}, \beta)^\mathsf{T}$ are right semi-continuous at $s = 0$ and integrable in y with respect to the probability measure $\mu(A) = \int_A p(z, t \mid X(0), 0)\mathrm{d}z$. We further require the transition density to satisfy usual regularity conditions given in Proposition 11. Such regularity conditions can be verified by using prior estimates of solutions in the theory of partial differential equations, which is far above the scope of this thesis.

In Chapter 6, where we used the concept of ultimate efficiency to assess the quality of designs, we restricted ourselves to linear Itō stochastic differential equations. The key was the statement of Lemma 4. In general, evaluation of the Fisher information for a process driven by the more general stochastic differential equation (7.1) is computationally challenging, because set of instances with explicit closed form transition densities is limited (except Ornstein-Uhlenbeck processes this includes, for instance, the Feller square root process [27] described by $\mathrm{d}X(t) = \theta_1[\theta_2 - X(t)]\mathrm{d}t + \beta\sqrt{X(t)}\mathrm{d}W(t)$).

Besides the naive discretisation method also known as Euler method, which might not perform ver well [65, 58], some authors developed more advanced approximate (pseudo-likelihood) techniques for maximum likelihood estimation (proxies for densities or likelihood functions directly), of which we highlight Elerian's method [21] and Aït-Sahalia's method [1, 2]. While the approximate methods give reasonable results for estimation, they might be sensitive to the guess at the true value of the unknown parameters; see, e. g., [58].

Consequently, for computation of the Fisher information matrix for processes governed by (7.1) we have to usually use greedy numerical procedures for solving the Kolmogorov's forward equation and for computation of its derivatives with respect to the unknown vector parameter at a prescribed point. This gives a strong argument for the use of the ultimate

efficiency in deciding whether we need to employ a battery of (thanks to nonlinearity and nonconvexity usually, again, greedy) optimisation methods.

In this chapter we outline further ideas in evaluating of the amount of information contained in a trajectory of a process driven by the stochastic differential equation (7.1).

7.1 Limiting behaviour of the solution to the Kolmogorov's equation

Let $p(t,x) = \frac{\partial}{\partial x}\Pr[X(t) < x]$ be a solution to the initial problem (1.10), and let us denote $\sigma^2(t,x)/2 = D(t,x)$, so that we use the conventional notation.

The main problem with the Kolmogorov's forward equation is the singularity at the initial time – the Dirac delta function. Using the infinitesimal generator [73] of the process described by (7.1), we can write the solution in the integral form as follows:

$$\frac{\partial p}{\partial t}(t,x) = \int_{\xi\in\mathbb{R}} \left(f(t,\xi)\frac{\partial}{\partial x} + D(t,\xi)\frac{\partial^2}{\partial x^2} \right) \delta(x-\xi)p(t,\xi)\mathrm{d}\xi.$$

By setting the well-known form for the Dirac delta function $\delta(x-\xi) = \int_{-\mathtt{i}\infty}^{\mathtt{i}\infty} \frac{1}{2\pi\mathtt{i}} \mathrm{e}^{z(\xi-x)}\mathrm{d}z$, where $\mathtt{i}$ denotes the imaginary unit, and by exploiting properties of exponential function, we get

$$p(t+\Delta,x) = \int_{\xi\in\mathbb{R}} \int_{-\mathtt{i}\infty}^{\mathtt{i}\infty} \frac{1}{2\pi\mathtt{i}} \exp\left\{ \Delta\left(-z\frac{x-\xi}{\Delta} + zf(t,\xi) + z^2 D(t,\xi) \right) \right\} p(t,\xi)\mathrm{d}z\mathrm{d}\xi + \mathrm{o}(\Delta).$$

Using the transform $w = z/\mathtt{i}$, some algebraic manipulations yield

$$p(t+\Delta,x) = \int_{\xi\in\mathbb{R}} \left(\int_{w\in\mathbb{R}} \frac{\sqrt{2D(t,\xi)\Delta}}{\sqrt{2\pi}} \exp\left\{ -\left[w + \mathtt{i}\frac{\frac{x-\xi}{\Delta} - f(t,\xi)}{2D(t,\xi)} \right]^2 D(t,\xi)\Delta \right\} \mathrm{d}w \right) \times \frac{1}{\sqrt{2\pi[2D(t,\xi)\Delta]}} \exp\left\{ -\frac{[x-(\xi+f(t,\xi)\Delta)]^2}{2[2D(t,\xi)\Delta]} \right\} p(t,\xi)\mathrm{d}\xi + \mathrm{o}(\Delta).$$

The integral $\int_{w\in\mathbb{R}} \ldots \mathrm{d}w$ is equal to 1, hence, by setting $2D(t,\xi) = \sigma^2(t,\xi)$ we can write

$$\begin{aligned} p(t+\Delta,x) &= \int_{\xi\in\mathbb{R}} \Psi\left[x;\xi+f(t,\xi)\Delta;\sigma^2(t,\xi)\Delta\right] p(t,\xi)\mathrm{d}\xi + \mathrm{o}(\Delta) \\ &= \int_{\xi\in\mathbb{R}} \left(\Psi\left[x;\xi+f(t,\xi)\Delta;\sigma^2(t,\xi)\Delta\right] + \mathrm{o}(\Delta) \right) p(t,\xi)\mathrm{d}\xi, \end{aligned}$$

where $\Psi[\cdot;\mu;\upsilon^2]$ is the probability density function of a Gaussian random variable with the mean μ and variance υ^2.

Consequently, we obtain

$$p(x,\Delta \mid y,t) = \Psi\left[x;y+f(t,y)\Delta;\sigma^2(t,y)\Delta\right] + \mathrm{o}(\Delta). \tag{7.2}$$

7.2 Ultimate information

Due to assumptions on regularity of transition density of the process given by equation (7.1) and semi-continuity of its derivatives, we obtain that

$$\mathscr{I}_{X(t+\Delta)|X(t)}(\boldsymbol{\vartheta}^*) = \left.\frac{\frac{\partial f(t,X(t))}{\partial \boldsymbol{\vartheta}} \frac{\partial f(t,X(t))}{\partial \boldsymbol{\vartheta}^\intercal}}{\sigma^2(t,X(t))}\right|_{\boldsymbol{\vartheta}^*} \Delta + \mathrm{o}(\Delta) + \frac{1}{2}\left.\frac{\partial}{\partial \boldsymbol{\vartheta}} \ln \sigma^2(t,X(t)) \frac{\partial}{\partial \boldsymbol{\vartheta}^\intercal} \ln \sigma^2(t,X(t))\right|_{\boldsymbol{\vartheta}^*}.$$

Conjecture 1. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by stochastic differential equation* (7.1) *satisfying the assumptions given thereafter, and let $\{\boldsymbol{\tau}^{(n)}\}_{n\geq m} \in \mathfrak{C}_{\mathcal{D}}$. Then*

$$\lim_{n\to\infty} \left(\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*) - \mathscr{I}_\infty(\boldsymbol{\vartheta}^*) - \sum_{i=2}^{n} \mathscr{O}(t_i, \boldsymbol{\vartheta}^*) \right) = \mathbf{0}_{m\times m},$$

where

$$\mathscr{I}_\infty(\boldsymbol{\vartheta}^*) = \mathscr{I}_{X(t_1)|X(0)}(\boldsymbol{\vartheta}) + \int_{T_*}^{T^*} \mathcal{E}_{X(t)} \left[\left. \frac{\frac{\partial f_{\boldsymbol{\vartheta}}(t,X(t))}{\partial \boldsymbol{\vartheta}} \frac{\partial f_{\boldsymbol{\vartheta}}(t,X(t))}{\partial \boldsymbol{\vartheta}^\intercal}}{\sigma^2_\beta(t,X(t))} \right|_{\boldsymbol{\vartheta}^*} \right] \mathrm{d}t,$$

and

$$\mathscr{O}(t,\boldsymbol{\vartheta}^*) = \frac{1}{2}\mathcal{E}_{X(t)} \left[\left. \frac{\partial}{\partial \boldsymbol{\vartheta}} \ln \sigma^2_\beta(t,X(t)) \frac{\partial}{\partial \boldsymbol{\vartheta}^\intercal} \ln \sigma^2_\beta(t,X(t)) \right|_{\boldsymbol{\vartheta}^*} \right].$$

Note that by setting $f(t,x) = a(t) + b(t)x$ and $\sigma(t,x) = \sigma(t)$, Theorem 1 directly yields the result of Theorem 7.

In more common situations, the difficulty is, again, that it is not easy to find the distribution of $X(t)$ explicitly. Nevertheless, given the prior guess $\boldsymbol{\vartheta}^*$ at the true value of the unknown parameter, numerical evaluation of the ultimate information in Theorem 1 is not complicated.

Chapter 8

Relation of the results with recent publications

In this chapter we point out the relation between the results of Chapter 6 and recent publications.

8.1 Concerning the existence of optimal designs

In Chapter 3 we explained that the property of "present simple" defined by Sack and Ylvisaker [92] yields continuity of the Fisher information matrix. Continuity of the Fisher information matrix is the sufficient condition in general, since, using the technique of Lemma 4, we can write

$$\mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_1),\boldsymbol{X}(\boldsymbol{\tau}_2)} = \mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_1)} + \mathcal{E}_{\boldsymbol{X}(\boldsymbol{\tau}_1)}[\mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_2)|\boldsymbol{X}(\boldsymbol{\tau}_1)}] \succeq_{\mathrm{L}} \mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_1)},$$

and if $\boldsymbol{\tau}_2 \to \boldsymbol{\tau}_0 \subseteq \boldsymbol{\tau}_1$, then the continuity gives a convergence of $\mathcal{E}_{\boldsymbol{X}(\boldsymbol{\tau}_1)}[\mathscr{I}_{\boldsymbol{X}(\boldsymbol{\tau}_2)|\boldsymbol{X}(\boldsymbol{\tau}_1)}]$ to zero matrix. What Sacks and Ylvisaker did was that rewrote the continuity condition into the language of functional spaces, which fitted their theory based on the reproducing kernel Hilbert spaces.

Chapter 6 revealed that for a stochastic differential equation (6.1), all the elements of the corresponding Fisher information matrix are continuous except the diagonal entry linked to the nuisance parameter β. Consequently, Theorem 5 and the discussion thereafter clarify the existence of optimal designs in the papers of Harman and Štulajter [39, 40], Harman [37] and Lacko [59], who took the parameter common for mean and variance as known. We can, however, move even further to stationary processes...

Consider the non-stationary Ornstein-Uhlenbeck process $\{X(t)\}_{t\geq 0}$ described by the well known equation $\mathrm{d}X(t) = \theta_1(\theta_2 - X(t)) + \beta \mathrm{d}W(t)$, that is, a process with the drift $f(t,x) = f(x) = \theta_1(\theta_2 - x)$ and constant volatility $\sigma(t) \equiv \beta$. If the value of parameter β is known, according to Theorem 5, a locally optimal design exists regardless of the choice of information function. Note that the lower boundary T_* of the experimental domain is not specified (only positivity is required), so the same properties hold when $T_* \to \infty$, i.e., when $\{X(t)\}_{t\geq 0}$ is a stationary Ornstein-Uhlenbeck process, where, for any $t_0 < t$, we have $\mathcal{E}[X(t)] = \theta_2$, $\mathcal{E}[X(t) \mid X(t_0)] = \mathrm{e}^{-\theta_1(t-t_0)}X(t_0) + (1 - \mathrm{e}^{-\theta_1(t-t_0)})\theta_2$, $\mathcal{V}[X(t)] = \beta^2/2\theta_1$, $\mathcal{V}[X(t) \mid X(t_0)] = \beta^2(1 - \mathrm{e}^{-2\theta_1(t-t_0)})/2\theta_1$ and $\mathcal{C}[X(t), X(t_0)] = \beta^2\mathrm{e}^{-2\theta_1(t-t_0)}/2\theta_1$. Actually, the stationary Ornstein-Uhlenbeck process is again a Gaussian Markov process; henceforth,

by using Lemma 4 in combination with the idea of Lemma 7, it is not hard to prove the continuity of the Fisher information matrix at the boundary points. Consequently, for the stationary Ornstein-Uhlenbeck process, if β is known, then there exists an optimal sampling design.

Now, let β be in the role of an unknown parameter of the non-stationary Ornstein-Uhlenbeck process. Obviously, $\{\mathscr{I}_\infty(t,\boldsymbol{\vartheta}^*)\}_{\beta\beta} = 0$, $\{\mathscr{I}_\infty(t,\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\theta}\beta} = \{\mathscr{I}_\infty(t,\boldsymbol{\vartheta}^*)\}_{\beta\boldsymbol{\theta}}^\mathsf{T} = \mathbf{0}_2$, and thus the matrix $\mathcal{Z}(\boldsymbol{\tau}_0,\boldsymbol{\alpha},\boldsymbol{\vartheta}^*)$ defined in Theorem 6 is singular for any $\boldsymbol{\alpha}$, $\boldsymbol{\vartheta}^*$ and, what is important, for any $\boldsymbol{\tau}_0$, which indicates potential problems with the existence of optimal sampling designs. Now, we investigate the behaviour of the Fisher information matrix for some $T_* \leq t_1 < t_d < t < t_u \leq T^*$ when t converges to t_d or t_u. It follows from Lemma 7 that we need to show the matrix inequality

$$\begin{aligned}\mathcal{E}_{X(t_d)}[\mathscr{I}_{X(t)|X(t_d)}(\boldsymbol{\vartheta}^*)] + \mathcal{E}_{X(t)}[\mathscr{I}_{X(t_u)|X(t)}(\boldsymbol{\vartheta}^*)] \\ - \mathcal{E}_{X(t_d)}[\mathscr{I}_{X(t_u)|X(t_d)}(\boldsymbol{\vartheta}^*)] - \begin{pmatrix}\mathbf{0}_{3\times3} & \mathbf{0}_3 \\ \mathbf{0}_3^\mathsf{T} & \frac{1}{2}\end{pmatrix} \max_{t\in\{t_d,t_u\}} \left(\frac{\partial \ln\sigma^2(t)}{\partial\beta}\right)^2 \succeq_\mathrm{L} \mathbf{0}_{4\times4} \quad (8.1)\end{aligned}$$

for some t. Since $\mathcal{V}[X(t+s) \mid X(t)] = \frac{\beta^2}{2\theta_1}(1-\mathrm{e}^{-2\theta_1 s})$, after some algebraic manipulation, we can show that $(\frac{\partial}{\partial\beta}\ln\mathcal{V}[X(t) \mid X(t_d)])^2 = (\frac{\partial}{\partial\beta}\ln\mathcal{V}[X(t_u) \mid X(t)])^2 = (\frac{\partial}{\partial\beta}\ln\sigma^2(t))^2 = 2/\beta$. Henceforth, the diagonal entry of the matrix obtained on the left-hand side of the inequality (8.1) that corresponds to β is zero, and if the matrix is non-negative definite, then the last column and the last row also have to be zero vectors for some t between t_d and t_u. From the fact that β is not present in the expectation of $X(t)$ (cf. (6.4)), it is not hard to find a necessary condition for the existence of a locally optimal design in the strong sense

$$\frac{\partial}{\partial\theta_1}(\ln\mathcal{V}[X(t) \mid X(t_d)] + \ln\mathcal{V}[X(t_u) \mid X(t)] - \ln\mathcal{V}[X(t_u) \mid X(t_d)]) = 0,\ t_d < t < t_u,$$

which is satisfied if and only if t equals t_d or t_u for any t_d and t_u (which are not feasible). Consequently, for the non-stationary Ornstein-Uhlenbeck process (6.3), the optimal design in the case of estimating $(X_0,\theta_1,\theta_2,\beta)^\mathsf{T}$ does not exist in the stronger sense. Additionally, the value of T_* was not specified, and thus, for $T_* \to \infty$, the same result also holds true for the stationary Ornstein-Uhlenbeck process. We can summarise that if the volatility parameter of the stationary and nonstationary Ornstein-Uhlenbeck processes is unknown, one has to verify the existence of an optimal sampling design for each selected optimality criterion separately.

The discussion of existence of optimal designs given above gives a complementary view of the results of Zagoraiou and Antognini [110], whose assumption that $\beta^2/2\theta_1$ stands for a single parameter leads to the situation when an optimal design for the estimation of θ_1 does not exist (although their Fisher information matrix is not proper). The restriction $\beta^2/2\theta_1$ causes a transition of the discontinuity of the Fisher information matrix in σ^2 also to θ_1. The focal point here is that for continuous processes, the properties of optimal designs are sensitive to the choice of parametrisation, especially if unknown parameters appear in the covariance structure.

8.2 Information matrices

Similarly to the form of the Fisher information matrix of Harman and Štulajter [39] for the processes with product covariance structures, Lemma 4 writes the Fisher information matrix for the Markov processes in an additive form with each summand depending on

two subsequent sampling times. Of course, both forms of the Fisher information matrices cover different areas of application.

Nonetheless, Lemma 4 might be useful also in instances with product covariance structure, especially when the unknown parameters are situated also in the covariance function. A typical example is the non-stationary Ornstein-Uhlenbeck process $\mathrm{d}X(t) = \theta_1(\theta_2 - X(t)) + \beta \mathrm{d}W(t)$, where the covariance structure is of the product form. If θ_2 and σ^2 are the only unknown parameters, then the use of the formula of Harman and Štulajter is straightforward, cf. [39]. On the other hand, when θ_1 is an unknown parameter too, then the Fisher information matrix is more complicated, where Lemma 4 yields more comfortable formulas. After some effort, we are able to write the Fisher information matrix for the non-stationary Ornstein-Uhlenbeck process in terms of the first sampling time and differences between sampling times, which is very popular for stationary processes; see, e.g., [53].

Resumé

Úvod

Mnohé javy, ktorých vlastnosti sú predmetom experimentálneho bádania, majú dynamický, evolučný charakter. V mnohých vedných oblastiach ako fyzika, biológia, medicína, inžinierstvo ale aj v spoločenskovedných disciplínach môžeme charakter správania sa skúmanej veličiny popísať diferenciálnymi rovnicami.

Napriek tomu, že charakterizácia mnohých fundamentálnych zákonov dobre aproximuje realitu, teoretické a pozorované skutočné hodnoty sa vždy odlišujú, pričom túto odlišnosť pripisujeme nepozorovateľným náhodným chybám. Preto neznáme vlastnosti (parametre diferenciálnych rovníc) odhadujeme.

V štatistike existujú dva principiálne odlišné spôsoby ako uchopiť náhodnosť pozorovaných hodnôt. Nech $\mathcal{L}$ je zákon popisujúci stav veličiny $x(t) = x_{\boldsymbol{\theta}}(t)$ v čase t v závislosti od hodnoty vektora vlastností $\boldsymbol{\theta}$. V prvom, klasickom prístupe predpokladáme, že veličina $x(t)$ spĺňa predpísaný zákon presne a výsledné pozorovania $X(t_i)$ v časoch t_i, $i = 1, \ldots, n$, sú potom výsledkom kontaminácie presných hodnôt $x(t_i)$ zväčša bielym šumom pripisovanému chybe meracieho zariadenia, čiže

$$\begin{cases} \mathcal{L}\left(\frac{\mathrm{d}^k x(t)}{\mathrm{d}t^k}, \frac{\mathrm{d}^{k-1} x(t)}{\mathrm{d}t^{k-1}}, \ldots, \frac{\mathrm{d}x(t)}{\mathrm{d}t}, x(t), t, \boldsymbol{\theta}\right) = 0, \\ X(t_i) = x(t_i) + \varepsilon_i, \ i = 1, \ldots, n. \end{cases} \tag{8.2}$$

Uvedený prístup možno nájsť v mnohých tradičných publikáciách.

Druhý prístup integruje šum priamo do zákona tým, že predpokladá

$$\mathcal{L}\left(\frac{\mathrm{d}^k X(t)}{\mathrm{d}t^k}, \frac{\mathrm{d}^{k-1} X(t)}{\mathrm{d}t^{k-1}}, \ldots, \frac{\mathrm{d}X(t)}{\mathrm{d}t}, X(t), t, \boldsymbol{\theta}, \text{“šum”}\right) = 0. \tag{8.3}$$

Na rigoróznu formuláciu prístupu (8.3) využívame terminológiu stochastického počtu, viď napríklad monografiu [73].

Základný rozdiel medzi situáciou popísanou v (8.2) a (8.3) spočíva v tom, že druhý prístup generuje stochastický proces s korelovanými pozorovaniami, ktorý pozorujeme presne, kým prvý prípad potláča prirodzenú vnútornú náhodnosť. Samozrejme, aj v prvom prístupe je možné uvažovať koreláciu medzi pozorovaniami, táto by však bola umelá.

Ciele dizertačnej práce

Predložená dizertačná práca sa zaoberá vývojom metód navrhovania experimentov pre procesy popísané Itōovymi stochastickými diferenciálnymi rovnicami (slabé riešenia) v tvare

$$\mathrm{d}X(t) = f_{\boldsymbol{\theta},\beta}(t, X(t))\mathrm{d}t + \sigma_\beta(t, X(t))\mathrm{d}W(t), \ X(0) = X_0 \tag{8.4}$$

kde $\boldsymbol{\theta}$ je vektor charakteristík procesu, ktorý je predmetom skúmania, β je parameter šumu mimo sféry záujmu, X_0 je neznáma ale pevná počiatočná hodnota a $\{W(t)\}_{t\geq 0}$ je Wienerov proces.

Stochastická diferenciálna rovnica (8.4) je pomerne všeobecná a uvedená oblasť nebola doposiaľ systematicky analyzovaná a dostupná literatúra ponúka iba zopár publikácií zaoberajúcich sa problematikou navrhovania experimentov pre procesy popísa(teľ)né stochastickými diferenciálnymi rovnicami, napríklad [92, 53, 110, 39, 40, 37, 3, 26]. preto sa v predloženej práci obmedzíme na jednoduchšie, lineárne rovnice v tvare

$$\mathrm{d}X(t) = [a_{\boldsymbol{\theta},\beta}(t) + b_{\boldsymbol{\theta},\beta}X(t)]\mathrm{d}t + \sigma_\beta(t)\mathrm{d}W(t), \tag{8.5}$$

čiže $f(t,x) = a(t) + b(t)x$ a $\sigma(t,x) = \sigma(t)$.

Cieľom dizertačnej práce je ...

- **... študovať existenciu optimálnych návrhov experimentov pre procesy popísané rovnicami v tvare** (8.5): Dôkaz existencie optimálnej alokácie časov pozorovania procesu je fundamentálnou otázkou, nakoľko z nej vyplýva opodstatnenosť použitia optimalizačných metód ako aj zmysel samotnej úlohy optimalizácie experimentu. Najčastejším spôsobom, ako zodpovedať uvedenú otázku, je stanovenie podmienok spojitosti Fisherovej informačnej matice na hranici množiny prípustných návrhov $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$; napríklad Sacks a Ylvisaker [92, 93] formulovali spojitosť Fisherovej informačnej matice cez spojitosť funkcionálnych priestorov, pričom sa odvolávali na vlastnosti Hilbertových priestorov s reprodukčnými jadrami. Vo väčšine publikácií, kde optimálne návrhy existujú (napr. [53, 39, 40, 37] a iné), autori uvažujú iba neznáme parametre strednej hodnoty. V prípadoch, kde je neznámy parameter situovaný aj v kovariančnej štruktúre, dochádza už ku komplikáciám, viď [110].

- **... odvodiť asymptotickú Fisherovu informačnú maticu**: Vypočítať optimálne časy pozorovania procesu s korelovanými pozorovaniami pri konečnom výbere je nekonvexný problém, a explicitné riešenia uvádza iba málo publikácií, napríklad [16, 53, 39, 40, 37]. Iní autori buď vyvíjali numerické metódy identifikácie optimálnych alokácií [81], alebo sa snažili obísť náročnosť uvedenej úlohy budovaním alternatívnych, často asymptotických (v zmysle počtu pozorovaní) metód [92, 93, 94, 106, 107, 24, 111, 17]. V dizertačnej práci tiež obchádzame základnú optimalizačnú úlohu, a to tak, že sa zameriavame na výpočet tzv. ultimátnej efektívnosti [79] (nomenklatúra [37]), kde porovnávame informáciu získanú z konečného návrhu s informáciou, ktorú by sme teoreticky získali pozorovaním celej trajektórie procesu. Ultimátna informácia má v plánovaní experimentov dvojaké využitie: na jednej strane nám poskytuje obraz o tom, do akej miery vyčerpávame totálnu dosiahnuteľnú informáciu a či má vôbec význam optimalizovať experiment, a na strane druhej nám umožňuje určiť, či náklady na ďalšie pozorovanie sú adekvátne eventuálnemu nárastu v množstve získanej informácie [92].

- **... rozšíriť výsledky získané pre lineárne stochastické diferenciálne rovnice v tvare** (8.5) **na všeobecnejšie stochastické diferenciálne rovnice v tvare** (8.4).

Motivácia: analýza neautonómneho nestacionárneho Ornsteinovho-Uhlenbeckovho procesu

Uvažujme Itōov proces $\{X(t)\}_{t\geq 0}$ popísaný stochastickou diferenciálnou rovnicou

$$\mathrm{d}X(t) = \kappa(\bar{X} - X(t))\mathrm{d}t + \sigma(t)\mathrm{d}W(t), \tag{8.6}$$

kde počiatočný stav $X(0) = X_0$ a stacionárna stredná hodnota $\bar{X}$ sú neznáme parametre, $\kappa > 0$ je známa rýchlosť konvergencie k stacionárnej strednej hodnote a $\sigma(t)$ je až na konštantný násobok známa polospojitá deterministická funkcia.

Použitím Itōovej lemy ľahko ukážeme, že

$$\mathcal{E}[X(t)] = \mathrm{e}^{-\kappa t} X_0 + (1 - \mathrm{e}^{-\kappa t})\bar{X},$$

čiže stredná hodnota procesu je lineárnou funkciou neznámych parametrov. Ak teda máme návrh $\boldsymbol{\tau} \in \overline{\mathfrak{T}}_n$ o rozsahu n, potom pozorovania $\boldsymbol{X}(\boldsymbol{\tau})$ spĺňajú lineárny regresný model

$$\boldsymbol{X}(\boldsymbol{\tau}) = (\mathbf{e}^{-\kappa\boldsymbol{\tau}})X_0 + (\mathbf{1}_n - \mathbf{e}^{-\kappa\boldsymbol{\tau}})\bar{X} + \boldsymbol{\varepsilon}(\boldsymbol{\tau}) = \mathbf{F}(\boldsymbol{\tau})\boldsymbol{\theta} + \boldsymbol{\varepsilon}(\boldsymbol{\tau}), \tag{8.7}$$

kde $\mathbf{F}(\boldsymbol{\tau}) = (\mathbf{e}^{-\kappa\boldsymbol{\tau}}, \mathbf{1}_n - \mathbf{e}^{-\kappa\boldsymbol{\tau}})$ je matica návrhu, $\boldsymbol{\theta} = (X_0, \bar{X})^\mathsf{T}$ je vektor neznámych parametrov a $\boldsymbol{\varepsilon}(\boldsymbol{\tau}) = (\varepsilon_{t_1}, \ldots, \varepsilon_{t_n})^\mathsf{T}$ je vektor náhodných chýb taký, že

$$\mathcal{E}[\boldsymbol{\varepsilon}(\boldsymbol{\tau})] = \mathbf{0}_n \text{ and } \mathcal{V}[\boldsymbol{\varepsilon}(\boldsymbol{\tau})] = \boldsymbol{\Sigma}(\boldsymbol{\tau}). \tag{8.8}$$

Keďže proces $\{X(t)\}_{t\geq 0}$ je gaussovský, pre výpočet informačnej matice potrebujeme charakterizovať aj druhé momenty pozorovaní.

Lema 1. *Majme vektor pozorovaní $\boldsymbol{X}(\boldsymbol{\tau})$, $\boldsymbol{\tau} \in \overline{\mathfrak{T}}_n$, potom ij-tý prvok, $i \leq j$, kovariančnej matice $\boldsymbol{\Sigma}(\boldsymbol{\tau})$ definovanej v* (8.8) *má tvar*

$$\begin{aligned} \{\boldsymbol{\Sigma}(\boldsymbol{\tau})\}_{ij} &= u(t_i)v(t_j), \text{ kde} \\ u(t) &= \mathrm{e}^{-\kappa t}\int_0^t \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu \text{ a} \\ v(t) &= \mathrm{e}^{-\kappa t}. \end{aligned}$$

Lema 2. *Kovariančná matica $\boldsymbol{\Sigma}(\boldsymbol{\tau})$ definovaná v leme 1 je kladne definitná pre každé $\boldsymbol{\tau} \in \mathfrak{T}_n$, $n \geq 2$.*

Z lemy 1 vyplýva, že proces popísaný rovnicou (8.6) má súčinovú kovariančnú štruktúru. Použitím výsledku Harmana a Štulajtera [39] o súčinových kovariančných štruktúrach dostaneme Fisherovu informačnú maticu pre návrh s rozsahom n:

$$\mathscr{I}(\boldsymbol{\tau}) = \begin{pmatrix} \frac{\mathrm{e}^{-2\kappa t_1}}{\mathcal{V}[X(t_1)]} & \frac{\mathrm{e}^{-\kappa t_1}(1-\mathrm{e}^{-\kappa t_1})}{\mathcal{V}[X(t_1)]} \\ \frac{\mathrm{e}^{-\kappa t_1}(1-\mathrm{e}^{-\kappa t_1})}{\mathcal{V}[X(t_1)]} & \frac{(1-\mathrm{e}^{-\kappa t_1})^2}{\mathcal{V}[D(t_1)]} + S(\boldsymbol{\tau}) \end{pmatrix},$$

kde

$$S(\boldsymbol{\tau}) = \sum_{i=2}^{n} \frac{(\mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}})^2}{\mathrm{e}^{2\kappa t_i}\mathcal{V}[X(t_i)] - \mathrm{e}^{2\kappa t_{i-1}}\mathcal{V}[X(t_{i-1})]} = \sum_{i=2}^{n} \frac{(\mathrm{e}^{\kappa t_i} - \mathrm{e}^{\kappa t_{i-1}})^2}{\int_{t_{i-1}}^{t_i} \mathrm{e}^{2\kappa\nu}\sigma^2(\nu)\mathrm{d}\nu}.$$

V ďalšom ukážeme, že existuje optimálny návrh o rozsahu n z množiny $\mathfrak{T}_n$. V prvom kroku ukazujeme, že optimálny návrh nie je koncentrovaný do jedného bodu.

Lema 3. *Nech $n \geq 3$ a $\boldsymbol{\tau}_0 = t_n\mathbf{1}_n$. Potom existuje návrh $\boldsymbol{\tau}_1 = (t_1, \ldots, t_n)^\mathsf{T} \in \overline{\mathfrak{T}}_n$ taký, že $t_1 < t_n$ a $\mathscr{I}(\boldsymbol{\tau}_1) \succeq_\mathrm{L} \mathscr{I}(\boldsymbol{\tau}_0)$.*

Keďže optimálny návrh je alokovaný vždy v aspoň dvoch odlišných bodoch, na dôkaz existencie optimálnych návrhov v množine $\mathfrak{T}_n$ stačí ukázať, že opakovaným pozorovaním nezískame žiadnu dodatočnú informáciu.

Veta 1. *V lineárnom modeli so strednou hodnotou* (8.7) *a kovariančnou štruktúrou uvedenou v leme 1 vždy existuje* Φ*-optimálny návrh o rozsahu* n*, na množine* $\mathfrak{T}_n$*.*

Veta 2. *V lineárnom modeli so strednou hodnotou* (8.7) *a kovariančnou štruktúrou uvedenou v leme 1, ak je funkcia* $\sigma(t)$ *nerastúca, potom je optimálne položiť* $t_n^* = T^*$*.*

Veta 3. *Asymptotická Fisherova informačná matica získaná pozorovaním procesu popísaného rovnicou* (8.6) *v každom čase na* $\mathcal{D}$ *je*

$$\mathscr{I}_\infty(T_*, T^*) = \begin{pmatrix} \frac{\mathrm{e}^{-2\kappa T_*}}{\mathcal{V}[X(T_*)]} & \frac{\mathrm{e}^{-\kappa T_*}(1-\mathrm{e}^{-\kappa T_*})}{\mathcal{V}[X(T_*)]} \\ \frac{\mathrm{e}^{-\kappa T_*}(1-\mathrm{e}^{-\kappa T_*})}{\mathcal{V}[X(T_*)]} & \frac{(1-\mathrm{e}^{-\kappa T_*})^2}{\mathcal{V}[X(T_*)]} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \kappa^2 \int_{T_*}^{T^*} \frac{\mathrm{d}\nu}{\sigma^2(\nu)}.$$

Navyše, pre ľubovoľný návrh $\boldsymbol{\tau} = (t_1, \ldots, t_n)^\mathsf{T} \in \mathfrak{T}_n$*, kde* $T_* \le t_1$ *a* $t_n \le T^*$*, platí:* ***i)*** $S(\boldsymbol{\tau}) \le S_\infty(T_*, T^*)$*, a* ***ii)*** $\mathscr{I}(\boldsymbol{\tau}) \preceq_\mathrm{L} \mathscr{I}_\infty(t_1, t_2) \preceq_\mathrm{L} \mathscr{I}_\infty(T_*, T^*)$*.*

Z vety 3 vyplýva, že optimálne časy pozorovania procesu sú viac koncentrované v oblastiach s nízkou úrovňou volatility, čo má zrejmú fyzikálnu interpretáciu, ktorú v práci uvádzame spolu s príkladom.

Výsledky pre všeobecné lineárne stochastické diferenciálne rovnice

Model a jeho vlastnosti

Predpokladajme, že pozorujeme proces $\{X(t)\}_{t\ge0}$ popísaný lineárnou Itōovou stochastickou diferenciálnou rovnicou v tvare

$$\begin{aligned} \mathrm{d}X(t) &= [a_{\boldsymbol{\theta},\beta}(t) + b_{\boldsymbol{\theta},\beta}(t)X(t)]\mathrm{d}t + \sigma_\beta(t)\mathrm{d}W(t) \\ &= f_{\boldsymbol{\theta},\beta}(t, X(t))\mathrm{d}t + \sigma_\beta(t)\mathrm{d}W(t), \\ X(0) &= X_0 \in \mathbb{R} \text{ je fixné.} \end{aligned} \tag{8.9}$$

Naše základné predpoklady týkajúce sa modelu sú: existujú derivácie funkcií $f_{\boldsymbol{\theta},\beta}(t,x)$ a $\sigma_\beta(t)$ vzhľadom na neznámy parameter $\boldsymbol{\vartheta} = (X_0, \boldsymbol{\theta}^\mathsf{T}, \beta)^\mathsf{T} \in \mathbb{R}\times\mathbb{R}^{m-2}\times\mathbb{R}$, funkcie $a_{\boldsymbol{\theta},\beta}(t)$, $\frac{\partial a_{\boldsymbol{\theta},\beta}(t)}{\partial\boldsymbol{\vartheta}}$, $b_{\boldsymbol{\theta},\beta}(t)$, $\frac{\partial b_{\boldsymbol{\theta},\beta}(t)}{\partial\boldsymbol{\vartheta}}$, $\sigma_\beta(t)$ a $\sigma_\beta^2(t)$ sú integrovateľné vzhľadom na t na intervale $[0, T^*]$, a $\sigma_\beta(t)$ je kladná takmer všade vzhľadom na Lebesgueovú mieru na reálnej priamke.

Kľúčom ku mnohým odpovediam a dôkazom nasledujúcich tvrdení, ktoré práca dáva, je fakt, že rovnica (8.9) generuje markovovský gaussovský proces. To nám umožní rozpísať Fisherovu informačnú maticu pre analýzy veľmi vhodným spôsobom:

Lema 4. *Nech* $\{X(t)\}_{t\ge0}$ *s pevným* $X(0)$ *je* $\boldsymbol{\vartheta}$*-parametrizovaný markovský proces so spojitým časom a gaussovskou prechodovou hustotou. Potom pre každý návrh* $\boldsymbol{\tau} \in \overline{\mathfrak{T}}_n$*, Fisherova informačná matica pre* $\boldsymbol{X}(\boldsymbol{\tau})$ *má tvar*

$$\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) = \mathscr{I}_{X(t_1)|X(0)}(\boldsymbol{\vartheta}^*) + \sum_{i=2}^{n} \mathcal{E}_{X(t_{i-1})}\left[\mathscr{I}_{X(t_i)|X(t_{i-1})}(\boldsymbol{\vartheta}^*)\right],$$

kde $\mathscr{I}_{X(t_i)|X(t_{i-1})}(\boldsymbol{\vartheta}^*)$ *je Fisherova informačná matica pre* $X(t_i)$ *podmienené hodnotou* $X(t_{i-1})$ *a* $\mathcal{E}_{X(t_{i-1})}[\cdot]$ *je očakávanie vzhľadom na náhodnú veličinu* $X(t_{i-1})$*.*

Lema 4 platí aj vo všeobecnejšej situácii, kedy prechodová hustota spĺňa tradičné podmienky regularity, a to aj v prípadoch, kedy pozorujeme viacrozmerný stochastický proces.

Základom pre hlbšiu analýzu procesu $\{X(t)\}_{t\geq 0}$ z hľadiska informácie sú jeho pravdepodobnostné vlastnosti. Zrejme v každom čase t môžeme sledovaný proces zapísať v tvare $X(t) \mid X(t_0) = \mathcal{E}[X(t) \mid X(t_0)] + \varepsilon(t) \mid X(t_0)$. Ak $B(t)$ je ľubovoľná primitívna funkcia k funkcii $b(t)$, potom

$$\begin{aligned}\mathcal{E}_{\boldsymbol{\theta},\beta}[X(t) \mid X(t_0)] &= \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(t_0)}X(t_0) + \int_{t_0}^{t} \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(\nu)}a_{\boldsymbol{\theta},\beta}(\nu)\mathrm{d}\nu,\\ \varepsilon(t) \mid X(t_0) &= \int_{t_0}^{t} \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(\nu)}\sigma_\beta(\nu)\mathrm{d}W(\nu)\end{aligned}$$

a z Itōovej izometrie vyplýva

$$\mathcal{V}_{\boldsymbol{\theta},\beta}[X(t) \mid X(t_0)] = \int_{t_0}^{t} \mathrm{e}^{2[B_{\boldsymbol{\theta},\beta}(t)-B_{\boldsymbol{\theta},\beta}(\nu)]}\sigma^2_\beta(\nu)\mathrm{d}\nu.$$

Aby sme si utvorili celkový obraz o $\{X(t)\}_{t\geq 0}$, dokázali sme nasledujúce pomocné tvrdenie:

Lema 5. *Nech $\{X(t)\}_{t\geq 0}$ je proces popísaný stochastickou diferenciálnou rovnicou* (8.9)*. Potom pre ľubovoľné t_1 a t_2, $t_2 \geq t_1 \geq 0$,*

$$\begin{aligned}\mathcal{C}_{\boldsymbol{\theta},\beta}[X(t_1), X(t_2)] &= u(t_1)v(t_2), \text{ kde}\\ u_{\boldsymbol{\theta},\beta}(t) &= \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)} \int_0^t \mathrm{e}^{-2B_{\boldsymbol{\theta},\beta}(\nu)}\sigma^2_\beta(\nu)\mathrm{d}\nu,\\ v_{\boldsymbol{\theta},\beta}(t) &= \mathrm{e}^{B_{\boldsymbol{\theta},\beta}(t)}.\end{aligned}$$

Čiže problém návrhu experimentu pre proces (8.9) sme pretransformovali na úlohu návrhu experimentu pre gaussovský nelineárny regresný model

$$\forall_{\boldsymbol{\tau}\in\overline{\mathfrak{T}}_n}\ \boldsymbol{X}(\boldsymbol{\tau}) \sim \mathcal{N}\Big(\mathcal{E}_{X_0,\boldsymbol{\theta},\beta}[\boldsymbol{X}(\boldsymbol{\tau})], \mathcal{V}_{\boldsymbol{\theta},\beta}[\boldsymbol{X}(\boldsymbol{\tau})]\Big)$$

Lema 6. *Nech $\{X(t)\}_{t\geq 0}$ je proces popísaný stochastickou diferenciálnou rovnicou* (8.9) *a $\|\cdot\|$ je metrika na $\mathbb{R}^n$. Pre každý návrh $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n$ a $\delta > 0$ existuje návrh $\boldsymbol{\tau} \in \mathfrak{T}_n$ taký, že $\|\boldsymbol{\tau}_0 - \boldsymbol{\tau}\| < \delta$ a $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) \succeq_{\mathrm{L}} \mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\vartheta}^*)$. Špeciálne, ak návrh $\boldsymbol{\tau}_0$ patrí do hraničnej množiny $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$, potom $\boldsymbol{\tau}_0$ je dominovaný ľubovoľným návrhom $\boldsymbol{\tau} \in \mathfrak{T}_n$, kde $t_1 = \{\boldsymbol{\tau}_0\}_1$ a $t_i = \{\boldsymbol{\tau}_0\}_i$ ak $\{\boldsymbol{\tau}_0\}_i > \{\boldsymbol{\tau}_0\}_{i-1}$.*

Dôkaz lemy 6 je založený na markovovskosti procesu $\{X(t)\}_{t\geq 0}$ a na skutočnosti, že $\lim_{\Delta\to 0}\Pr[X(t+\Delta) = X(t)] = 1$, a preto uvedený výsledok môžeme zovšeobecniť na širšiu triedu procesov, podobne ako v prípade lemy 4.

Keďže proces $\{X(t)\}$, ktorý rieši rovnicu (8.9), je gaussovský, Fisherova informačná matica má tvar

$$\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) = \begin{pmatrix} \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_1\boldsymbol{\vartheta}_1} & \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_1\beta} \\ \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\beta\boldsymbol{\vartheta}_1} & \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\beta\beta} \end{pmatrix},$$

kde

$$\begin{aligned}\{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2} = \Bigg(&\frac{\partial\mathcal{E}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial\boldsymbol{\alpha}_1}\mathcal{V}^{-1}[\boldsymbol{X}(\boldsymbol{\tau})]\frac{\partial\mathcal{E}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial\boldsymbol{\alpha}_2^{\mathsf{T}}}\Bigg)\Bigg|_{\boldsymbol{\vartheta}^*}\\ &+ \frac{1}{2}\mathrm{tr}\left\{\mathcal{V}^{-1}[\boldsymbol{X}(\boldsymbol{\tau})]\frac{\partial\mathcal{V}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial\boldsymbol{\alpha}_1}\mathcal{V}^{-1}[\boldsymbol{X}(\boldsymbol{\tau})]\frac{\partial\mathcal{V}[\boldsymbol{X}(\boldsymbol{\tau})]}{\partial\boldsymbol{\alpha}_2^{\mathsf{T}}}\right\}\Bigg|_{\boldsymbol{\vartheta}^*} \qquad (8.10)\end{aligned}$$

a $\boldsymbol{\vartheta}^*$ je prvotný odhad skutočnej hodnoty neznámeho parametra [67].

Veta 4. *Ak je počiatočná hodnota X_0 stochastickej diferenciálnej rovnice* (8.9) *jediným neznámym parametrom, potom je optimálne položiť $t_1 = T_*$ bez ohľadu na rozsah návrhu. Odpovedajúca variancia odhadu je*

$$\mathcal{V}[\hat{X}_0] = \int_0^{T_*} \mathrm{e}^{2[B(0)-B(\nu)]}\sigma^2(\nu)\mathrm{d}\nu.$$

Dôležitým predpokladom pre platnosť predchádzajúcej vety je absencia počiatočnej hodnoty v koeficientoch stochastickej diferenciálnej rovnice.

Existencia lokálne optimálnych návrhov

Ťažkosti s existenciou optimálnych návrhov, a to nielen pre uvažovanú triedu procesov, ale pre modely s korelovanými pozorovaniami vo všeobecnosti, vznikajú, keď Fisherova informačná matica $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ nie je spojitá v hraničných bodoch $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$. To zdôrazňuje dôležitosť asymptotických vlastností Fisherovej informačnej matice. Samozrejme, špecifikáciou informačnej funkcie Φ, t. j., miera informácie je definovaná hodnotou $\Phi[\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)]$, môžeme potlačiť vplyv nespojitosti Fisherovej informačnej matice, preto sa zaoberáme otázkou existencie optimálnych návrhov v "silnom zmysle", čiže sa pýtame, či pre každý bod z hraničnej množiny $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$ existuje návrh $\boldsymbol{\tau} \in \mathfrak{T}_n$ taký, že pre ľubovoľnú postupnosť návrhov $\{\boldsymbol{\tau}^{(k)}\}_k$ on $\mathfrak{T}_n$, kde $\lim_{k\to\infty} \boldsymbol{\tau}^{(k)} = \boldsymbol{\tau}_0$, Fisherova informačná matica $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ loewnerovsky dominuje maticu $\lim_{k\to\infty} \mathscr{I}(\boldsymbol{\tau}^{(k)}, \boldsymbol{\vartheta}^*)$. To znamená, že ak je splnená podmienka existencie optimálneho návrhu v silnom zmysle, potom existuje optimálny návrh bez ohľadu na výber informačnej funkcie.

Pomocou lemy 4 môžeme preskúmať limitné vlastnosti Fisherovej informačnej matice pre podkladovú stochastickú diferenciálnu rovnicu (8.9):

Lema 7. *Nech $\{X(t)\}_{t\geq 0}$ je proces popísaný stochastickou diferenciálnou rovnicou* (8.9) *a $\boldsymbol{\vartheta}^*$ je prvotný odhad skutočnej hodnoty parametra $\boldsymbol{\vartheta}$. Potom, pre ľubovoľné konštanty π_1 and π_2,*

$$\lim_{\Delta\to 0} \left(\mathcal{E}_{X(t+\pi_1\Delta)}\left[\mathscr{I}_{X(t+\pi_2\Delta)|X(t+\pi_1\Delta)}(\boldsymbol{\vartheta}^*)\right] - \mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)(\pi_2 - \pi_1)\Delta - \mathscr{O}(t, \boldsymbol{\vartheta}^*)\right) = \mathbf{0}_{m\times m},$$

kde

$$\mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*) = \left.\frac{\frac{\partial f_{\boldsymbol{\vartheta}}(t,x)}{\partial \boldsymbol{\vartheta}}\Big|_{x=\mathcal{E}[X(t)]} \frac{\partial f_{\boldsymbol{\vartheta}}(t,x)}{\partial \boldsymbol{\vartheta}^\mathsf{T}}\Big|_{x=\mathcal{E}[X(t)]} + \frac{\partial b_{\boldsymbol{\vartheta}}(t)}{\partial \boldsymbol{\vartheta}} \frac{\partial b_{\boldsymbol{\vartheta}}(t)}{\partial \boldsymbol{\vartheta}^\mathsf{T}} \mathcal{V}_{\boldsymbol{\vartheta}}[X(t)]}{\sigma^2_\beta(t)}\right|_{\boldsymbol{\vartheta}^*}$$

a

$$\mathscr{O}(t, \boldsymbol{\vartheta}^*) = \left.\frac{1}{2}\frac{\partial \ln \sigma^2_\beta(t)}{\partial \boldsymbol{\vartheta}} \frac{\partial \ln \sigma^2_\beta(t)}{\partial \boldsymbol{\vartheta}^\mathsf{T}}\right|_{\boldsymbol{\vartheta}^*}.$$

Keďže volatilita $\sigma(t)$ závisí výhradne na parametri β, podľa lemy 7 jediný prvok Fisherovej informačnej matice, ktorý potenciálne nie je spojitý na hraničnej množine, je diagonálny prvok $\{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\beta\beta}$. Preto platí nasledujúca veta:

Veta 5. *Ak β nie je neznámy parameter stochastickej diferenciálnej rovnice* (8.9)*, potom optimálny návrh pre odhad parametra $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_\mathrm{I}$ existuje v silnom zmysle, a to pre každé n a $\boldsymbol{\vartheta}^*$.*

Technika dôkazu vety 5 môže byť aplikovaná aj keď $\boldsymbol{\theta} = (\boldsymbol{\theta}_\mathcal{E}^\mathsf{T}, \boldsymbol{\theta}_\mathcal{V}^\mathsf{T})^\mathsf{T}$, kde $\frac{\partial}{\partial \boldsymbol{\theta}_\mathcal{V}}\mathcal{E}[X(t)] = \frac{\partial}{\partial \beta}\mathcal{E}[X(t)] = 0$ a $\frac{\partial}{\partial \boldsymbol{\theta}_\mathcal{E}}\mathcal{V}[X(t)] = \mathbf{0}$. Potom Fisherova informácia o $(X_0, \boldsymbol{\theta}_\mathcal{E}^\mathsf{T})^\mathsf{T}$ vyplývajúca zo Schurovho doplnku je rovná bloku Fisherovej informačnej matice $\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ zodpovedajúcej $(X_0, \boldsymbol{\theta}_\mathcal{E}^\mathsf{T})^\mathsf{T}$. Tento blok je však spojitý, a teda optimálny návrh pre odhad $(X_0, \boldsymbol{\theta}_\mathcal{E}^\mathsf{T})^\mathsf{T}$ existuje v silnom zmysle.

Vo všeobecnejšej situácii, za predpokladu korelácie pozorovaní je otázka existencie optimálnych návrhov komplexný problém, a to hlavne z dôvodu nekonvexnosti informácie.

Jednou z možností ako preskúmať, či sa supremálna informácia dosahuje pri konvergencii postupnosti návrhov k hraničnej množine, je zamerať sa na lokálne správanie sa Fisherovej informácie na hraničnej množine $\overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$. Na tento účel môžeme použiť pojem smerovej derivácie [77]. V oblasti navrhovania experimentov, pre konvexné množiny, pre ξ a ζ dané je smerová derivácia (maticovej/skalárnej) funkcie $\boldsymbol{h}$ v bode $\boldsymbol{\xi}$ a smere $\boldsymbol{\zeta}$ zvyčajne definovaná ako

$$\partial \boldsymbol{h}(\boldsymbol{\xi}, \boldsymbol{\zeta} - \boldsymbol{\xi}) = \lim_{\Delta \searrow 0} \frac{\boldsymbol{h}[\boldsymbol{\xi} + \Delta(\boldsymbol{\zeta} - \boldsymbol{\xi})] - \lim_{\Delta \searrow 0} \boldsymbol{h}[\boldsymbol{\xi} + \Delta(\boldsymbol{\zeta} - \boldsymbol{\xi})]}{\Delta}.$$

Uvedená smerová derivácia je dodefinovaná pre inštancie, kedy funkcia $\boldsymbol{h}$ nie je spojitá v bode ξ.

Predpoklady modelu však akdávajú určité obmedzenia na výber prípustných smerov. Pre ľubovoľný hraničný bod $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$ skonštruujeme prípustný smer $\boldsymbol{\tau} = (t_1, \ldots, t_n)^\mathsf{T}$ tak, že vezmeme $t_i = \{\boldsymbol{\tau}_0\}_i + \pi_i$, $i = 1, \ldots, n$, kde $\pi_i = 0$ if $\{\boldsymbol{\tau}_0\}_i > \{\boldsymbol{\tau}_0\}_{i-1}$, a vezmeme $\pi_i > \pi_{i-1} \geq 0$ ak $\{\boldsymbol{\tau}_0\}_i = \{\boldsymbol{\tau}_0\}_{i-1}$ a $t_i > t_{i-1}$. ďalej vyžadujeme, aby $\{\boldsymbol{\tau}_0\}_i + \pi_i > \{\boldsymbol{\tau}_0\}_{i-1} + \pi_{i-1}$. Pre Fisherovu informačnú maticu potom môžeme definovať

$$\partial \mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\pi}, \boldsymbol{\vartheta}^*) = \lim_{\Delta \searrow 0} \frac{\mathscr{I}(\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}, \boldsymbol{\vartheta}^*) - \lim_{\Delta \searrow 0} \mathscr{I}(\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}, \boldsymbol{\vartheta}^*)}{\Delta}.$$

Vektor $\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}$ má ostro rastúce zložky a konverguje k $\boldsymbol{\tau}_0$ pre $\Delta \to 0$. Navyše, smerová derivácia $\partial \mathscr{I}(\boldsymbol{\tau}_0, \boldsymbol{\pi}, \boldsymbol{\vartheta}^*)$ je pozitívne homogénna v $\boldsymbol{\pi}$, a preto podmienka $\{\boldsymbol{\tau}_0\}_i + \pi_i > \{\boldsymbol{\tau}_0\}_{i-1} + \pi_{i-1}$ nie je nevyhnutná, keďže pre ľubovoľne malé Δ sú zložky vektora $\boldsymbol{\tau}_0 + \Delta\boldsymbol{\pi}$ ostro rastúce. Postačuje teda slabšia podmienka na výber smeru $\boldsymbol{\pi}$:

$$\boldsymbol{\pi} = \begin{cases} 0, & \{\boldsymbol{\tau}_0\}_i > \{\boldsymbol{\tau}_0\}_{i-1} \\ \pi_i > \pi_{i-1}, & \{\boldsymbol{\tau}_0\}_i = \{\boldsymbol{\tau}_0\}_{i-1} \end{cases}.$$

Využitím modifikovanej smerovej derivácie sme dokázali nasledujúcu vetu:

Veta 6. *Nech $\{X(t)\}_{t \geq 0}$ je proces popísaný stochastickou diferenciálnou rovnicou* (8.9) *a $\boldsymbol{\vartheta}^*$ je prvotný odhad skutočnej hodnoty parametra $\boldsymbol{\vartheta}$. Ak pre hraničný návrh $\boldsymbol{\tau}_0 \in \overline{\mathfrak{T}}_n \setminus \mathfrak{T}_n$ existuje vektor $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\boldsymbol{\tau}_0)$ nezáporných konštánt taký, že matica*

$$\mathscr{Z}(\boldsymbol{\tau}_0, \boldsymbol{\alpha}, \boldsymbol{\vartheta}^*) = \sum_{i=1, \{\boldsymbol{\tau}_0\}_i = \{\boldsymbol{\tau}_0\}_{i-1}}^{n} \alpha_i \begin{pmatrix} \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\theta}\beta} \\ \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\beta\boldsymbol{\theta}} & \{\mathscr{I}_\infty(\{\boldsymbol{\tau}_0\}_i, \boldsymbol{\vartheta}^*)\}_{\beta\beta} \end{pmatrix}$$

je kladne definitná, potom sa supremálna informácia nedosahuje, keď konvergujeme k $\boldsymbol{\tau}_0$, a to bez ohľadu na voľbu informačnej funkcie.

Poznamenávame, že matica $\mathscr{I}_\infty(t, \boldsymbol{\vartheta}^*)$ má hodnosť najviac dva. Zo subaditivity hodnosti matice ako funkcie vyplýva, že hodnosť matice $\mathscr{Z}(\boldsymbol{\tau}_0, \boldsymbol{\alpha}, \boldsymbol{\vartheta}^*)$ je najviac dvojnásobok kardinality množiny $\{i : \{\boldsymbol{\tau}_0\}_i \neq \{\boldsymbol{\tau}_0\}_{i-1}\}$. Veta 6 je preto užitočná najmä v prípadoch, kedy sa overuje existencia optimálnych návrhov v silnom zmysle a $\dim(\boldsymbol{\theta}) = 1$ ($\dim(\boldsymbol{\vartheta}) = 3$). Napriek tomu môže táto veta dať určité indikácie aj pre vyššie dimenzie parametra $\boldsymbol{\theta}$.

Ultimátna efektívnosť návrhov

Definícia ultimátnej efektívnosti tak, ako je uvedená v článku [79] je aplikovateľná iba ak je $\lim_{n \to \infty} \Phi[\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)]$ konečná, čo v našej situácii, kde parameter β je konzistentne odhadnuteľný, nemusí platiť. Preto sme sa zamerali na subparameter $\boldsymbol{\vartheta}_\mathrm{I}$: Vo všeobecnosti, ak

$\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_{\mathrm{I}}^{\mathsf{T}}, \boldsymbol{\vartheta}_{\mathrm{II}}^{\mathsf{T}})^{\mathsf{T}}$ je delenie parametra na dva subparametre a $\boldsymbol{\tau} \in \mathfrak{T}_n$, potom Fisherova informačná matica $\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)$ zodpovedajúca $\boldsymbol{\vartheta}_{\mathrm{I}}$ je daná Schurovým doplnkom

$$\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*) = \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}} - \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}^{-}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{I}}},$$

kde $\{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}^{-}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}}$ je ľubovoľná zovšeobecnená inverzia matice $\{\mathscr{I}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}}$. Navyše Schurov doplnok zachováva loewnerovske usporiadanie matíc [63], preto:

Tvrdenie 1. *Ak optimálny návrh pre odhad parametra $\boldsymbol{\vartheta}$ existuje v silnom zmysle, potom optimálny návrh pre odhad parametra $\boldsymbol{\vartheta}_{\mathrm{I}}$ tiež existuje v silnom zmysle.*

Definícia 1 (Modifikácia definície ultimátnej informácie)**.** *Nech $\{\boldsymbol{\tau}^{(n)}\}_{n \geq m} \in \mathfrak{C}_{\mathcal{D}}$ a nech $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_{\mathrm{I}}^{\mathsf{T}}, \boldsymbol{\vartheta}_{\mathrm{II}}^{\mathsf{T}})^{\mathsf{T}}$ je rozdelenie neznámeho parametra, pre ktoré platí*

$$\lambda_{\max}\left(\{\mathscr{I}_{\infty}(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}\right) < \infty,$$

$\lambda_{\max}$ označuje najväčšie vlastné číslo, a

$$\{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}^{-}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{II}}} \{\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{II}}\boldsymbol{\vartheta}_{\mathrm{I}}} \to \mathbf{0}_{\dim(\boldsymbol{\vartheta}_{\mathrm{I}}) \times \dim(\boldsymbol{\vartheta}_{\mathrm{I}})}, \tag{8.11}$$

ako $n \to \infty$. Potom (lokálna) ultimátna efektívnosť návrhu $\boldsymbol{\tau} \in \mathfrak{T}_n$ pre odhad parametra $\boldsymbol{\vartheta}_{\mathrm{I}}$ vzhľadom na informačnú funkciu Φ, skrátene (lokálna) ultimátna Φ-efektívnosť, je pomer

$$\mathrm{ueff}(\boldsymbol{\tau} \mid \Phi, \boldsymbol{\vartheta}^*) = \frac{\Phi[\mathscr{I}_{\mathrm{I}}(\boldsymbol{\tau}, \boldsymbol{\vartheta}^*)]}{\Phi[\{\mathscr{I}_{\infty}(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\vartheta}_{\mathrm{I}}\boldsymbol{\vartheta}_{\mathrm{I}}}]}.$$

Z lemy 4 a lemy 7 sme dostali:

Veta 7. *Nech $\{X(t)\}_{t \geq 0}$ je proces popísaný stochastickou diferenciálnou rovnicou* (8.9) *a nech $\{\boldsymbol{\tau}^{(n)}\}_{n \geq m} \in \mathfrak{C}_{\mathcal{D}}$. Potom*

$$\lim_{n \to \infty} \left(\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*) - \mathscr{I}_{\infty}(\boldsymbol{\vartheta}^*) - \sum_{i=2}^{n} \mathscr{O}(t_i, \boldsymbol{\vartheta}^*) \right) = \mathbf{0}_{m \times m},$$

kde

$$\mathscr{I}_{\infty}(\boldsymbol{\vartheta}^*) = \left. \frac{\frac{\partial \mathcal{E}[X(T_*)]}{\partial \boldsymbol{\vartheta}} \frac{\partial \mathcal{E}[X(T_*)]}{\partial \boldsymbol{\vartheta}^{\mathsf{T}}}}{\mathcal{V}[X(T_*)]} \right|_{\boldsymbol{\vartheta}^*} + \frac{1}{2} \left. \frac{\partial \ln \mathcal{V}[X(T_*)]}{\partial \boldsymbol{\vartheta}} \frac{\partial \ln \mathcal{V}[X(T_*)]}{\partial \boldsymbol{\vartheta}^{\mathsf{T}}} \right|_{\boldsymbol{\vartheta}^*} + \int_{T_*}^{T^*} \mathscr{I}_{\infty}(t, \boldsymbol{\vartheta}^*) \mathrm{d}t,$$

a $\mathscr{I}_{\infty}(t, \boldsymbol{\vartheta}^*)$ a $\mathscr{O}(t, \boldsymbol{\vartheta}^*)$ sú definované v leme 7.

Keďže parameter β je konzistentne odhadnuteľný, použitím predchádzajúcej vety a v súlade s definíciou 1 vieme vypočítať ultimátnu Φ-efektívnosť návrhov pre subparameter $\boldsymbol{\vartheta}_{\mathrm{I}} = (X_0, \boldsymbol{\theta}^{\mathsf{T}})^{\mathsf{T}}$.

Procesy Ornsteinovho-Uhlenbeckovho typu

Predchádzajúce výsledky sa síce vzťahujú na lineárne stochastické diferenciálne rovnice, otázkou však je možnosť rozšírenia výsledkov aj na iné, nelineárne rovnice. Motiváciou tu je hlavne Fisherova-Neymannova veta o faktorizácii o postačujúcich štatistikách, ktorá nás doviedla k nasledujúcej definícii:

Definícia 2. *Nech $\mu_{\boldsymbol{\theta}\beta}(t,y)$ a $\gamma_\beta(t,y)$ sú dostatočne hladké funkciea nech proces $\{Y(t)\}_{t\geq 0}$ je popísaný stochastickou diferenciálnou rovnicou*

$$\mathrm{d}Y(t) = \mu_{\boldsymbol{\theta},\beta}(t,Y(t))\mathrm{d}t + \gamma_\beta(t,Y(t))\mathrm{d}W(t).$$

Ak existujú dostatočne hladké funkcie $a_{\boldsymbol{\theta},\beta}(t)$, $b_{\boldsymbol{\theta},\beta}(t)$, $\sigma_\beta(t)$ a $\varphi(t,y)$, kde φ je bijektívna v y a $\frac{\partial\varphi(t,y)}{\partial\boldsymbol{\vartheta}} = \mathbf{0}_m$ pre všetky $t \geq 0$ a $y \in \mathbb{R}$, také, že proces $\{X(t)\}_{t\geq 0} = \{\varphi(t,Y(t))\}_{t\geq 0}$ je popísaný rovnicou

$$\mathrm{d}X(t) = [a_{\boldsymbol{\theta},\beta}(t) + b_{\boldsymbol{\theta},\beta}(t)X(t)]\mathrm{d}t + \sigma_\beta(t)\mathrm{d}W(t),$$

potom hovoríme, že proces $\{Y(t)\}_{\geq 0}$ je Ornsteinovho-Uhlenbeckovho typu s asociovanými koeficientmi $a(t)$, $b(t)$ a $\sigma(t)$. Tento fakt označujeme symbolom $\{Y(t)\} \in \mathcal{OU}_{a(t),b(t),\sigma(t)}$.

Veta 8. *Nech $\{Y(t)\}_{t\geq 0}$ je proces popísaný stochastickou diferenciálnou rovnicou*

$$\mathrm{d}Y(t) = \mu(t,Y(t))\mathrm{d}t + \sigma(t)g(t,Y(t))\mathrm{d}W(t),$$

kde funkcie $\mu(t,y)$ and $g(t,y)$ sú dostatočne hladké a $\frac{\partial g(t,y)}{\partial\boldsymbol{\theta}} = 0$ pre všetky $t \geq 0$ a $y \in \mathbb{R}$. Ak je splnená podmienka

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\frac{\mathrm{d}y}{g(t,y)} = a(t) + b(t)\int\frac{\mathrm{d}y}{g(t,y)} + \frac{1}{2}\sigma^2(t)\frac{\partial g(t,y)}{\partial y} - \frac{\mu(t,y)}{g(t,y)}$$

pre nejaké funkcie $a_{\boldsymbol{\theta},\beta}(t)$ a $b_{\boldsymbol{\theta},\beta}(t)$, potom $\{Y(t)\}_{t\geq 0} \in \mathcal{OU}_{a_{\boldsymbol{\theta},\beta}(t),b_{\boldsymbol{\theta},\beta}(t),\sigma_\beta(t)}$.

Dôsledok 1. *Nech $\{Y(t)\}_{t\geq 0} \in \mathcal{OU}_{a_{\boldsymbol{\theta},\beta},b_{\boldsymbol{\theta},\beta},\sigma_\beta}$ je proces popísaný stochastickou diferenciálnou rovnicou*

$$\mathrm{d}Y(t) = \mu_{\boldsymbol{\theta},\beta}(Y(t))\mathrm{d}t + \sigma_\beta g(Y(t))\mathrm{d}W(t)$$

a nech $\mu(y)$ je dané. Potom $g(y)$ rieši obyčajnú diferenciálnu rovnicu

$$\left(b - \frac{\partial\mu(y)}{\partial y}\right)g(y) + \mu(y)\frac{\partial g(y)}{\partial y} + \frac{1}{2}\sigma^2 g^2(y)\frac{\partial^2 g(y)}{\partial y^2} = 0.$$

Príklad aplikácie výsledkov na Gompertzovom rastovom modeli

V Gompertzovom rastovom modeli [36] je očakávaná veľkosť nádoru $Y_{\theta_1,\theta_2}(t)$ riešením obyčajnej diferenciálnej rovnice

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = \theta_2 Y(t) - \theta_1 Y(t)\ln Y(t),\ Y(0) = Y_0 \text{ známe}, \tag{8.12}$$

kde faktor spomaľujúci rast θ_1 a vnútorná miera rastu θ_2 sú neznáme parametre v sfére záujmu, a počiatočná veľkosť Y_0 je zvyčajne známa a získaná z prvej detekcie (pri pokusoch z prvej infekcie). Riešenie rovnice (8.12) je známe v explicitnej uzavretej podobne:

$$Y_{\theta_1,\theta_2}(t) = \mathrm{e}^{\mathrm{e}^{-\theta_1 t}\ln Y_0 + (1-\mathrm{e}^{-\theta_1 t})\frac{\theta_2}{\theta_1}}, \tag{8.13}$$

čo vysvetľuje popularitu modelu v aplikáciách.

Lokálne D-optimálny návrh Gomperzovho modelu, kde očakávaná veľkosť nádoru (8.13) je kontaminovaná homoskedastickým bielym šumom, bol analyzovaný v článku [64].

Gompertzov model je možné prepísať do stochastickej diferenciálnej rovnice nasledovne [66]:

$$\mathrm{d}Y(t) = [\theta_2 Y(t) - \theta_1 Y(t)\ln Y(t)]\mathrm{d}t + \beta Y(t)\mathrm{d}W(t),\ Y(0) = Y_0 \text{ známe}. \tag{8.14}$$

Tu, čím väčší je nádor, tým kolísavejšia je jeho veľkosť a naopak.

Ak definujeme $\{X(t)\}_{t\geq 0} = \{\ln Y(t)\}_{t\geq 0}$, potom $\{X(t)\}_{t\geq 0}$ rieši stochastickú diferenciálnu rovnicu

$$\mathrm{d}X(t) = \left(\theta_2 - \frac{1}{2}\beta^2 - \theta_1 X(t)\right)\mathrm{d}t + \beta \mathrm{d}W(t).$$

Čiže $\{Y(t)\}_{t\geq 0}$ je proces Ornsteinovho-Uhlenbeckovho typu s asociovanými koeficientmi

$$a_{\theta_2,\beta}(t) = \theta_2 - \frac{1}{2}\beta^2,\ b_{\theta_1}(t) = -\theta_1 \text{ a } \sigma_\beta(t) = \beta.$$

Ak položíme $\boldsymbol{\theta} = (\theta_1, \theta_2)^\mathsf{T} = \boldsymbol{\vartheta}_\mathrm{I}$, dostaneme asymptotickú Fisherovu informačnú maticu pre $\boldsymbol{\theta}$

$$\{\mathscr{I}_\infty(\boldsymbol{\vartheta}^*)\}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\frac{\partial \mathcal{E}[X(T_*)]}{\partial \boldsymbol{\theta}} \frac{\partial \mathcal{E}[X(T_*)]}{\partial \boldsymbol{\theta}}}{\mathcal{V}[X(T_*)]} + \frac{\partial \ln \mathcal{V}[X(T_*)]}{\partial \boldsymbol{\theta}} \frac{\partial \ln \mathcal{V}[X(T_*)]}{\partial \boldsymbol{\theta}} + \frac{1}{\beta^2}\int_{T_*}^{T^*} \begin{pmatrix} \mathcal{E}^2[X(t)] + \mathcal{V}[X(t)] & -\mathcal{E}[X(t)] \\ -\mathcal{E}[X(t)] & 1 \end{pmatrix} \mathrm{d}t,$$

kde $\mathcal{E}[X(t)] = \mathrm{e}^{-\theta_1 t}\ln Y_0 + \frac{1}{\theta_1}(\theta_2 - \frac{1}{2}\beta^2)(1 - \mathrm{e}^{-\theta_1 t})$ a $\mathcal{V}[X(t)] = \frac{\beta^2}{2\theta_1}(1 - \mathrm{e}^{-2\theta_1 t})$.

Týmto sme získali všetky potrebné údaje pre výpočet ultimátnej efektívnosti návrhov.

Zovšeobecnenie výsledkov pre asymptotickú Fisherovu informačnú maticu

Uvažujme všeobecnú situáciu, kedy pozorujeme Itōov proces $\{X(t)\}_{t\geq 0}$ popísaný stochastickou diferenciálnou rovnicou

$$\mathrm{d}X(t) = f_{\boldsymbol{\theta},\beta}(t, X(t))\mathrm{d}t + \sigma_\beta(t, X(t))\mathrm{d}W(t). \tag{8.15}$$

Predpokladáme, že existuje slabé riešenie rovnice (8.15), jeho prechodová hustota $p(x, s \mid y, t) = \frac{\partial}{\partial x}\Pr[X(t+s) < x \mid X(t) = y]$ a jej prvé a druhé derivácie vzhľadom na neznámy parameter $\boldsymbol{\vartheta} = (\boldsymbol{\theta}^\mathsf{T}, \beta)^\mathsf{T}$ sú polospojité sprava v bode $s = 0$ a integrovateľné v y vzhľadom na pravdepodobnostnú mieru $\mu(A) = \int_A p(z, t \mid X(0), 0)\mathrm{d}z$. ďalej požadujeme, aby prechodová hustota spĺňala obvyklé podmienky regularity.

Potom dostávame nasledujúcu hypotĂ©zu:

Hypotéza 1. *Nech $\{X(t)\}_{t\geq 0}$ je proces popísaný stochastickou diferenciálnou rovnicou* (8.15) *spĺňajúcou príslušné predpoklady a nech $\{\boldsymbol{\tau}^{(n)}\}_{n\geq m} \in \mathfrak{C}_\mathcal{D}$. Potom*

$$\lim_{n\to\infty}\left(\mathscr{I}(\boldsymbol{\tau}^{(n)}, \boldsymbol{\vartheta}^*) - \mathscr{I}_\infty(\boldsymbol{\vartheta}^*) - \sum_{i=2}^{n}\mathscr{O}(t_i, \boldsymbol{\vartheta}^*)\right) = \mathbf{0}_{m\times m},$$

kde

$$\mathscr{I}_\infty(\boldsymbol{\vartheta}^*) = \mathscr{I}_{X(t_1)|X(0)}(\boldsymbol{\vartheta}) + \int_{T_*}^{T^*}\mathcal{E}_{X(t)}\left[\left.\frac{\frac{\partial f_{\boldsymbol{\vartheta}}(t,X(t))}{\partial \boldsymbol{\vartheta}}\frac{\partial f_{\boldsymbol{\vartheta}}(t,X(t))}{\partial \boldsymbol{\vartheta}^\mathsf{T}}}{\sigma^2_\beta(t, X(t))}\right|_{\boldsymbol{\vartheta}^*}\right]\mathrm{d}t,$$

a

$$\mathscr{O}(t, \boldsymbol{\vartheta}^*) = \frac{1}{2}\mathcal{E}_{X(t)}\left[\left.\frac{\partial}{\partial \boldsymbol{\vartheta}}\ln\sigma^2_\beta(t, X(t))\frac{\partial}{\partial \boldsymbol{\vartheta}^\mathsf{T}}\ln\sigma^2_\beta(t, X(t))\right|_{\boldsymbol{\vartheta}^*}\right].$$

Bibliography

[1] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70:223–262, 2002.

[2] Y. Aït-Sahalia. Closed-form likelihood expansions for a multivariate diffusions. *The Annals of Statistics*, 36:906–937, 2008.

[3] V. V. Anisimov, V. V. Fedorov, and S. L. Leonov. Optimal design of pharmacokinetic studies described by stochastic differential equations. *mODa 8 – Advances in Model-Oriented Design and Analysis*, pages 9–16, 2007.

[4] L. Arnold. *Stochastic Differential Equations: Theory And Applications*. John Wiley & Sons, 1974.

[5] L. Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 3:21–86, 1900.

[6] P. J. Bickel and A. M. Herzberg. Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *The Annals of Statistics*, 7:77–95, 1979.

[7] P. J. Bickel, A. M. Herzberg, and M. F. Schilling. Robustness of Design Against Autocorrelation in Time II: Optimality, Theoretical and Numerical Results for the First-Order Autoregressive Process. *Journal of the American Statistical Association*, 76:870–877, 1981.

[8] R. Brown. A Brief Account of Microscopical Observations Made in the Months of June, July, and August 1827, on the Particles contained in the Pollen of Plants; and on the General Existence of active Molecules in Organic and Inorganic Bodies. *Edinburgh New Philosophical Journal*, 5:358–371, 1828.

[9] W. Bryc. *The Normal Distribution: Characterization with Applications (Lecture Notes in Statistics)*. Springer, 1995.

[10] D. A. Cameron. Mathematical Modelling of the Response of Breast Cancer to Drug Therapy. *Journal of Theoretical Medicine*, 1:137–151, 1997.

[11] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, Thomson Learning, 2nd edition, 2002.

[12] H. Chernoff. Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24:585–602, 1953.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[14] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[15] M. J. Crowder. Maximum Likelihood Estimation for Dependent Observations. *Journal of the Royal Statistical Society. Series B*, 38:45–53, 1976.

[16] H. Dette, J. Kunert, and J. Pepelyshev. Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica*, 18:135–154, 2008.

[17] H. Dette, A. Pepelyshev, and A. Zhigljavsky. Optimal design for linear models with correlated observations. *The Annals of Statistics*, 41:143–176, 2013.

[18] R. Durrett. *Stochastic Calculus*. CRC Press, 1996.

[19] W. Ebeling, E. Gudowska-Nowak, and I. M. Sokolov. On stochastic dynamics in physics – remarks on history and terminology. *Acta Physica Polonica B*, 39:1003–1018, 2008.

[20] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 15:549–560, 1905.

[21] O. Elerian. A note on the existence of a closed form conditional density for the Milstein scheme. *Working paper, Nuffield College, Oxford University*, 1998.

[22] G. Elfving. Convex sets in statistics. In *Tolfte Skandinaviska Matematikerkongressen*, pages 34–39, 1953.

[23] R. L. Eubank, P. L. Smith, and P. W. Smith. Uniqueness and eventual uniqueness of optimal designs in some time series models. *The Annals of Statistics*, 9:486–493, 1981.

[24] R. L. Eubank, P. L. Smith, and P. W. Smith. A note on optimal and asymptotically optimal designs for certain time series models. *The Annals of Statistics*, 10:1295–1301, 1982.

[25] R. L. Eubank, P. L. Smith, and P. W. Smith. On the computation of optimal designs for certain time series models with applications to optimal quantile selection for location or scale parameter estimation. *SIAM Journal on Scientific and Statistical Computing*, 3:238–249, 1982.

[26] V. V. Fedorov, S. L. Leonov, and V. A. Vasiliev. Pharmacokinetic Studies Described by Stochastic Differential Equations: Optimal Design for Systems with Positive Trajectories. *mODa 9 – Advances in Model-Oriented Design and Analysis*, pages 73–80, 2010.

[27] W. Feller. Two Singular Diffusion Problems. *The Annals of Mathematics*, 54:173–182, 1951.

[28] J. Fellman. On the allocation of linear observations. *Scientiarum Fennica Commentationes Physico-Mathematicae*, 44:27–78, 1974.

[29] R. A. Fisher. The logic of inductive inference (with discussion). *Journal of the Royal Statistical Society*, 98:39–82, 1935.

[30] A. D. Fokker. Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld. *Annalen der Physik*, 248:810–820, 1914.

[31] B. R. Frieden. *Science from Fisher Information: A Unification*. Cambridge University Press, 2nd edition, 2004.

[32] A. Friedman. *Stochastic Differential Equations and Applications*. Academic Press, 1975.

[33] R. Gagnon and S Leonov. Optimal population designs for PK models with serial sampling. *Journal of Biopharmaceutical Statistics*, 15:143–163, 2005.

[34] F. Gantmacher. *The Theory of Matrices*. Chelsea Publishing Company, 1959.

[35] C. W. Gardiner. *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences*. Springer, 1985.

[36] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–585, 1825.

[37] R. Harman. On exact optimal sampling designs for processes with a product covariance structure. In *Experiments for Processes with Time or Space Dynamics*, Isaac Newton Institute, Cambridge, 2011.

[38] R. Harman and T. Jurík. Computing c-optimal designs using the simplex method of linear programming. *Computational Statistics & Data Analysis*, 53:247–254, 2008.

[39] R. Harman and F. Štulajter. Optimality of equidistant sampling designs for a nonstationary Ornstein-Uhlenbeck process. *Proceedings of the 6th St. Petersburg Workshop on Simulation*, 2:1097–1101, 2009.

[40] R. Harman and F. Štulajter. Optimal sampling designs for the Brownian motion with a quadratic drift. *Journal of Statistical Planning and Inference*, 141:2480–2488, 2011.

[41] P. G. Hoel. Efficiency Problems in Polynomial Estimation. *The Annals of Mathematical Statistics*, 29:1134–1145, 1958.

[42] P. G. Hoel. Asymptotic Efficiency in Polynomial Estimation. *The Annals of Mathematical Statistics*, 32:1042–1047, 1961.

[43] M. Hofmanová and J. Seidler. On weak solutions of stochastic differential equations. *Stochastic Analysis and Applications*, 30:100–121, 2012.

[44] R. A. Horn and Ch. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[45] S. M. Iacus. *Simulation and inference for stochastic differential equations*. Springer, 2008.

[46] J. Ingen-Housz. *Vermischte schriften physisch-medicinischen inhalts*. Wappler, 1784.

[47] K. Itō. On a formula concerning stochastic differentials. *Nagoya Mathematical Journal*, 3:55–65, 1951.

[48] J. A. Jacquez. *Compartmental Analysis in Biology and Medicine*. Elsevier, 1972.

[49] S. Karlin. *Total Positivity*. Stanford University Press, 1968.

[50] J. Kiefer. Optimum experimental designs. *Journal of the Royal Statistical Society*, 21:273–319, 1959.

[51] J. Kiefer. Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models*, pages 333–353, 1974.

[52] J. Kiefer. General equivalence theory for optimum designs. *The Annals of Statistics*, 2:849–879, 1974.

[53] J. Kiseľák and M. Stehlík. Equidistant and D-optimal designs for parameters of Ornstein-Uhlenbeck process. *Statistics & Probability Letters*, 78:1388–1396, 2008.

[54] F. B. Knight. *Essentials of Brownian Motion and Diffusion*. American Mathematical Society, 1981.

[55] A. Kolmogorov. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104:415–458, 1931.

[56] A. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Julius Springer, 1933.

[57] Y. K. Kwok. *Mathematical Models of Financial Derivatives*. Springer, 1998.

[58] V. Lacko. Two-factor Convergence Model of Cox-Ingersoll-Ross Type. Master's thesis, Comenius University, 2010.

[59] V. Lacko. Planning of experiments for a nonautonomous Ornstein-Uhlenbeck process. *Tatra Mountains Mathematical Publications*, 51:101–113, 2012.

[60] V. Lacko. Ultimate efficiency of experimental designs for Ornstein-Uhlenbeck type processes. *to appear in Journal of Statistical Planning and Inference*, 2014.

[61] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, 2nd edition, 1998.

[62] D. S. Lemons. *An introduction to stochastic processes in physics*. John Hopkins UP, 2002.

[63] C.-K. Li and R. Mathias. Extremal characterizations of the Schur Complement and resulting inequalities. *SIAM Review*, 42:233–236, 2000.

[64] G. Li. Optimal and efficient designs for Gompertz regression models. *Annals of Institute of Statistical Mathematics*, 64:945–957, 2012.

[65] A. Lo. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Economic Theory*, 4:231–247, 1988.

[66] C. F. Lo. Stochastic Nonlinear Gompertz Model of Tumour Growth. In *Proceedings of the World Congress on Engineering*, 2009.

[67] K. V. Mardia and R. T. Marshall. Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika*, 71:135–146, 1984.

[68] L. Michaelis and M. L. Menten. Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift: Beiträge zur chem. Physiologie u. Pathologie*, 49:333–369, 1913.

[69] B. Mukherjee. Exactly optimal sampling designs for processes with a product covariance structure. *Canadian Journal of Statistic*, 31:69–87, 2003.

[70] W. Näther. *Effective Observation of Random Fields. Teubner-Texte zur Mathematik* **72**. Teubner, 1985.

[71] W. Näther. Exact designs for regression models with correlated errors. *Statistics*, 16:479–484, 1985.

[72] L. Norton. A Gompertzian Model of Human Breast Cancer Growth. *Cancer Research*, 44:7067–7071, 1988.

[73] B. Øksendal. *Stochastic Differential Equations*. Springer, 2000.

[74] E. Parzen. An approach to time series analysis. *The Annals of Mathematical Statistics*, 32:951–989, 1961.

[75] E. Parzen. Regression analysis of continuous parameter time series. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:469–489, 1961.

[76] A. Pázman. Application of basic relations of adjustment computation for time-continuous measurings. *Acta Metronomica*, 1, 1965.

[77] A. Pázman. *Foundations of optimum experimental design*. Riedel, 1986.

[78] A. Pázman. *Nonlinear statistical models*. Kluwer, 1993.

[79] A. Pázman. Criteria for optimal design of small-sample experiments with correlated observations. *Kybernetika*, 43:453–462, 2007.

[80] A. Pázman. Information Contained in Design Points of Experiments with Correlated Observations. *Kybernetika*, 46:771–783, 2010.

[81] A. Pázman and W. G. Müller. Optimal design of experiments subject to correlated errors. *Statistics & Probability Letters*, 52:29–34, 2001.

[82] M. Planck. Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 24:323–341, 1917.

[83] J. W. Pratt. F. Y. Edgeworth and R. A. Fisher on the Efficiency of Maximum Likelihood Estimation. *The Annals of Statistics*, 4:501–514, 1976.

[84] L. Pronzato and Pázman A. *Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Springer, 2013.

[85] F. Pukelsheim. On linear regression designs which maximize information. *Journal of Statistical Planning and Inference*, 4:339–364, 1980.

[86] F. Pukelsheim. Efficient rounding of approximate designs. *Biometrika*, 79:763–770, 1992.

[87] F. Pukelsheim. *Optimal design of experiments*. Wiley, 1993.

[88] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

[89] C. R. Rao. *Linear Statistical Inference and its Applications*. John Wiley & Sons, 1973.

[90] L. M. Riccardi and L. Sacerdote. The Ornstein-Uhlenbeck process as a model for neuronal activity. *Biological Cybernetics*, 37:1–9, 1979.

[91] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer, 1996.

[92] J. Sacks and D. Ylvisaker. Designs for regression problems with correlated errors. *The Annals of Mathematical Statistics*, 37:66–89, 1966.

[93] J. Sacks and D. Ylvisaker. Designs for regression problems with correlated errors: Many parameters. *The Annals of Mathematical Statistics*, 39:49–69, 1968.

[94] J. Sacks and D. Ylvisaker. Designs for regression problems with correlated errors III. *The Annals of Mathematical Statistics*, 41:2057–2074, 1970.

[95] G. Sagnol. Computing Optimal Designs of multiresponse Experiments reduces to Second-Order Cone Programming. *Journal of Statistical Planning and Inference*, 141:1684–1708, 2010.

[96] Z. Schuss. *Theory and applications of stochastic processes: An analytical approach*. Springer, 2010.

[97] G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. John Wiley & Sons, 1989.

[98] A. V. Skorokhod. On existence and uniqueness of solutions to stochastic diffusion equations. *Sibirskii Matematicheskii Zhurnal*, 2:129–137, 1961.

[99] A. V. Skorokhod. On stochastic differential equations. In *Proceedings of the 6th all-union conference on probability theory and mathematical statistics*, pages 159–168, 1962.

[100] J. F. Speer, V. E. Petrosky, M. W. Retsky, and et al. A Stochastic Numerical Model of Breast Cancer Growth That Simulates Clinical Data. *Cancer Research*, 44:4124–4130, 1984.

[101] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer, 1979.

[102] P. W. Sullivan and S. E. Salmon. Kinetics of Tumor Growth and Regression in IgG Multiple Myeloma. *The Journal of Clinical Investigation*, 51:1697–1708, 1972.

[103] T. J. Sweeting. Uniform Asymptotic Normality of the Maximum Likelihood Estimator. *The Annals of Statistics*, 8:1375–1381, 1980.

[104] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the Brownian motion. *Physical Review*, 36:823–841, 1930.

[105] L. Vandenberghe and S. Boyd. Applications of semidefinite programming. *Applied Numerical Mathematics*, 29:283–299, 1999.

[106] G. Wahba. On the regression design problem of Sacks and Ylvisaker. *The Annals of Mathematical Statistics*, 42:1035–1053, 1971.

[107] G. Wahba. Regression design for some equivalence classes of kernels. *The Annals of Statistics*, 2:925–934, 1974.

[108] N. Wiener. Differential space. *Journal of Mathematical Physics*, 2:131–174, 1923.

[109] Y. Yu. Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics*, 38:1593–1606, 2010.

[110] M. Zagoraiou and A. B. Antognini. Optimal design for parameter estimation of the Ornstein-Uhlenbeck process. *Applied Stochastic Models in Business and Industry*, 25:583–600, 2009.

[111] A. Zhigljavsky, H. Dette, and A. Pepelyshev. A new approach to optimal design for linear models with correlated observations. *Journal of American Statistical Association*, 105:1093–1103, 2010.