COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

ULTIMATE EFFICIENCY OF DESIGNS FOR MULTIVARIATE ORNSTEIN-UHLENBECK PROCESSES

MASTER'S THESIS

2014 Bc. Michal Hojčka

COMENIUS UNIVERSITY IN BRATISLAVA
Faculty of Mathematics, Physics and Informatics
Department of Applied Mathematics and Statistics

ULTIMATE EFFICIENCY OF DESIGNS FOR MULTIVARIATE ORNSTEIN-UHLENBECK PROCESSES

Master's Thesis

Bc. Michal HOJČKA

9.1.9 Applied Mathematics
Economic and Financial Mathematics

Supervisor:
RNDr. Vladimír LACKO

BRATISLAVA 2014

Ultimate efficiency of designs for multivariate Ornstein-Uhlenbeck processes

Bc. Michal Hojčka
E-mail: *michal.hojcka@gmail.com*

RNDr. Vladimír Lacko
E-mail: *lackovladimir@gmail.com*

Department of Applied Mathematics and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
Mlynská dolina
842 48 Bratislava
Slovak Republic

Master's Thesis in 9.1.9 Applied Mathematics
Compilation date: April 28, 2014
Typeset in LaTeX

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

68167823

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Michal Hojčka
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Ultimate efficiency of designs for multivariate Ornstein-Uhlenbeck processes
Ultimátna efektívnosť návrhov pre viacrozmerné Ornstein-Uhlenbeckove procesy

Cieľ: Študovať návrh experimentov pre procesy s korelovanými pozorovaniami popísané pomocou stochastických diferenciálnych rovníc. Posudzovať kvalitu návrhov pre viacrozmerné Ornstein-Uhlenbeckove procesy pomocou asymptotickej Fisherovej informačnej matice.

Vedúci: RNDr. Vladimír Lacko
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Dátum zadania: 25.01.2013

Dátum schválenia: 04.02.2013 prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Abstract

In the master's thesis we focus on finding optimal design of experiments for the processes described by stochastic differential equations. In the first part we outline the theory on stochastic processes and finding optimal design for processes with correlated observations. In the second part we present our stochastic model for multivariate Ornstein-Uhlenbeck processes and we discuss existence of optimal design for finding the unknown parameter. As a reference to measure the quality of design we use asymptotic Fisher information matrix.

Keywords: Multivariate Ornstein-Uhlenbeck process • Fisher information matrix • Optimal design • Ultimate efficiency.

Acknowledgement

I would like to express my gratitude to my supervisor RNDr. Vladimír Lacko for help and time he gave me throughout the process of writing the thesis. I would also like to thank my family and friends for support.

Declaration on Word of Honour

I hereby declare this thesis was written on my own with using referred literature and help of my supervisor.

.....

Bc. Michal Hojčka

UNIVERZITA KOMENSKÉHO V BRATISLAVE
Fakulta matematiky, fyziky a informatiky
Katedra aplikovanej matematiky a štatistiky

ULTIMÁTNA EFEKTÍVNOSŤ NÁVRHOV PRE VIACROZMERNÉ ORNSTEIN-UHLENBECKOVE PROCESY

Diplomová práca

Bc. Michal HOJČKA

9.1.9 aplikovaná matematika
Ekonomická a finančná matematika

Školiteľ:
RNDr. Vladimír LACKO

BRATISLAVA 2014

Abstrakt

Diplomová práca sa zaoberá hľadaním optimálneho návrhu experimentov pre procesy popísané pomocou stochastických diferenciálnych rovníc. V prvej časti zhrnieme základy teórie stochastických procesov a optimálneho návrhu experimentov pre procesy s korelovanými pozorovaniami. V druhej časti predstavíme model popisujúci viacrozmerné Ornstein-Uhlenbeckove procesy a zaoberáme sa existenciou optimálneho návrhu pre odhad parametrov daného modelu. Na určovanie kvality návrhov ako referenčnú hodnotu používame asymptotickú Fisherovu informačnú maticu.

Kľúčové slová: Viacrozmerný Ornstein-Uhlenbeckov proces • Fisherova informačná matica • Optimálny návrh • Ultimátna efektivita.

Contents

Introduction

After some time, it became my habit, even for finite dimensional probabilistic phenomena, to look at an infinite dimensional set-up, the properties of which may illuminate those of the finite dimensional set-up considered previously.

Kiyoshi Itō

Almost all processes in the real world can be described by stochastic differential equations, as randomness is a part of everything. Thanks to Kiyoshi Itō and his contribution in [3], the theory on solving these equations is on quite advanced level. We can meet with stochastic processes in physics, medicine, finances and many other fields of interest. One of the most famous and oldest models in applications is Gompertz model of tumor growth [1]. Problem that arises from finding optimal design of experiments for stochastic processes is the fact that observations within the single process are correlated and therefore it does not help to repeat the observation twice in the same time. Solution could be to run the process again, but it is often unreal, since many parameters of processes in pharmacy like tumor growth depends individually on each patient. From the origin of experimental designs in 1747, when surgeon James Lind carried out the first controlled experiment to develop a cure for scurvy up the present, medicine is still a field, where design of experiments can apply as performing medical scans is very costly and each patient can usually undergo only a few examinations. Unfortunately, most of these experiments are not optimised from the mathematic point of view. As researches suggest, even with only a few observations we can get quite high efficiency of given design. Saks and Ylvisaker[12] used asymptotic efficiency to decide, whether performing another observation is worth its new information gain.

In the present thesis we focus on multivariate Ornstein-Uhlenbeck processes, which covers wide class of real-life processes. Their origins lies in the work of Ornstein and Uhlenbeck [11], where they studied velocity of the particle. Instead of determining the optimal design, we focus mainly on existence of optimal design in this thesis, trying to decide for which parameters of the processes exists such design. Later we measure the quality of designs through ultimate efficiency, method originated in [9] and [2]. Key reference in this case is asymptotic Fisher information matrix, constructed from the observation of the whole trajectory of the process. We employ the methods suggested by Lacko in [5] for univariate processes and try to extend it to multivariate case to find practical way to compute this matrix and thus determite the situations in which ultimate efficiency has reasonable application.

CHAPTER 1

Introduction to stochastic calculus

In this chapter we provide some basic knowledge about stochastic calculus. We refer the reader to the monograph of Øksendal [8], which is the base for the presented chapter.

Definition 1. *A stochastic process is a parametrized collection of random variables* $\{X_t\}_{t\in T}$ *defined on a probability space* $(\Omega, \mathcal{F}, P)$ *and assuming values in* $\mathbb{R}^n$.

The parameter space T is usually the halfline $[0, \infty)$, but it may also be interval $[a, b]$, the non-negative integers and even subsets of $\mathbb{R}^n$ for $n \geq 1$. For each $t \in T$ fixed we have a random variable $\omega \mapsto X_t(\omega)$; $\omega \in \Omega$. On the other hand, fixing $\omega \in \Omega$ we can consider the function $t \mapsto X(\omega)$; $t \in T$ which is called a path of X_t. We usually think of t as "time" and each ω as individual "particle" or "experiment". Thus we may also regard the process as a function of two variables $(t, \omega) \mapsto X(t, \omega)$ from $T \times \Omega \mapsto \mathbb{R}^n$. The (finite dimensional) distributions of the process $\{X_t\}_{t\in T}$ are the measures $\mu_{t_1,\cdots,t_k}$ defined on $\mathbb{R}^{nk}$ by $\mu_{t_1,\cdots,t_k}(F_1 \times \ldots \times F_k) = \Pr[X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k]$; $t_i \in T$.

Theorem 2 (Kolmogorov Extension Theorem)**.** *For all* $t_1, \cdots, t_k \in T, k \in \mathbb{N}$*, let* $\nu_{t_1,\cdots,t_k}$ *be probability measures on* $\mathbb{R}^{nk}$ *such that*

$$\nu_{t_{\sigma(1)},\cdots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = \nu_{t_1,\cdots,t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)}) \qquad (K1)$$

for all permutations σ *on* $\{1, 2, \cdots, k\}$ *and*

$$\nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\cdots,t_k,t_{k+1},\cdots t_{k+m}}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)} \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n) \quad (K2)$$

for all $m \in \mathbb{N}$*, where the set on the right hand side has a total of* $k + m$ *factors. Then there exists a probability space* $(\Omega, \mathcal{F}, P)$ *and a stochastic process* $\{X_t\}$ *on* Ω*,* $X_t : \Omega \mapsto \mathbb{R}^n$*, such that* $\nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = P[X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k]$ *for all* $t_i \in T$*,* $k \in \mathbb{N}$ *and all Borel sets* F_i.

Definition 3. *A filtration on* $(\Omega, \mathcal{F})$ *is a family* $\mathcal{M} = \{\mathcal{M}_t\}_{t\geq 0}$ *of* σ*-algebras* $\mathcal{M}_t \subset \mathcal{F}$ *such that*

$$0 \leq s < t \Rightarrow \mathcal{M}_s \subset \mathcal{M}_t$$

(i.e. $\{\mathcal{M}_t\}$ *is increasing).*

Let us now define the most basic, but very important example of a stochastic process.

Definition 4 (Wiener process)**.** *A one-dimensional Wiener process is a process, defined by the following three properties:*

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

ii) For any partition $t_0, \ldots, t_n$*, the increments* $W(t_n) - W(t_{n-1}), W(t_{n-1}) - W(t_{n-2}), \ldots, W(t_1) - W(t_0)$ *are independent.*

iii) For any $t, s \geq 0$*,* $W(t+s) - W(t) \sim \mathcal{N}(0, s)$.

An n*-dimensional Wiener process is a process of which components are independent one-dimensional Wiener processes.*

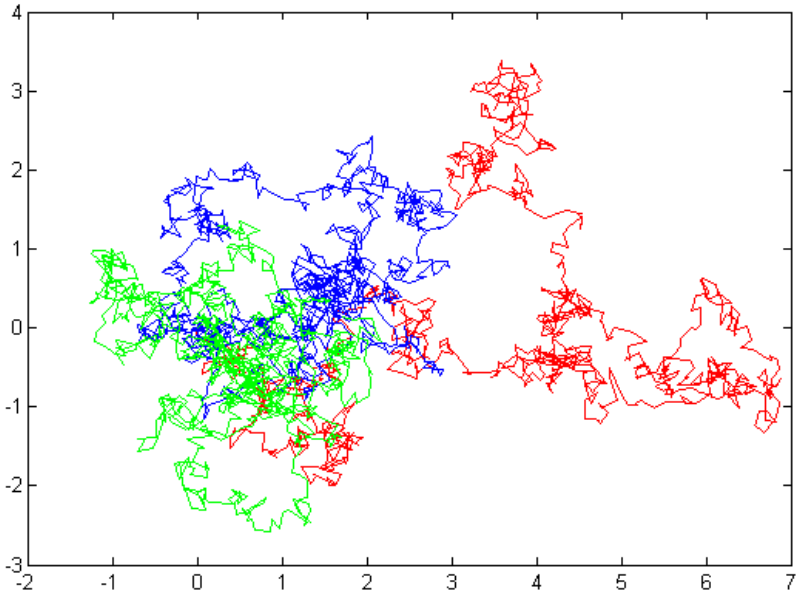

Figure 1.1: *Examples of* $2-$*dimensional Wiener processes*

Definition 5. *Suppose that $\{X_t\}$ and $\{Y_t\}$ are stochastic processes on $(\Omega, \mathcal{F}, P)$. Then we say that $\{X_t\}$ is a version of $\{Y_t\}$ if $\Pr[\{\omega; X_t(\omega) = Y_t(\omega)\}] = 1$ for all t. Note that if X_t is a version of Y_t, then X_t and Y_t have the same finite-dimensional distributions. Thus from the point of view that a stochastic process is a probability law on $(\mathbb{R}^n)^{[0,\infty)}$ two such processes are the same, but nevertheless their path properties may be different.*

Theorem 6 (Kolmogorov's continuity theorem)**.** *Suppose that the process $X = \{X_t\}_{t\geq 0}$ satisfies the following condition: For all $T > 0$ there exist positive constants α, β, D such that $E[|X_t - X_s|^\alpha] \leq D \cdot |t - s|^{1+\beta}$; $0 \leq s, t \leq T$. Then there exists a continuous version of X.*

As $n-$dimensional Wiener process satisfies Kolmogorov's condition with $\alpha = 4$, $D = n(n + 2)$ and $\beta = 1$, it follows that it has a continuous version. From now on we will assume that W_t is such a continuous version.

Definition 7. *If $X_t(.) : \Omega \mapsto \mathbb{R}$ is a continuous stochastic process, then quadratic variation of process X_t, $\langle X, X\rangle_t^{(2)}$ is defined by*

$$\langle X, X\rangle_t^{(2)}(\omega) = \lim_{\Delta t_k \to 0} \sum_{t_k \leq t} |X_{t_{k+1}}(\omega) - X_{t_k}(\omega)|^2 \quad \textit{(limit in probability)}$$

where $0 = t_1 < t_2 < \cdots < t_n = t$ and $\Delta t_k = t_{k+1} - t_k$.

From the properties of Wiener process we obtain that quadratic variation of Wiener process is finite and for the course of the process during time t it achieves also value t.

Definition 8. *Let $W_t(\omega)$ be n-dimensional Wiener process. Then we define $\mathcal{F}_t = \mathcal{F}_t^{(n)}$ to be the σ-algebra generated by the random variables $W_s(.); s \leq t$. We assume that all sets of measure zero are included in $\mathcal{F}_t$. A function $h(\omega)$ will be $\mathcal{F}_t$-measurable if and only if h can be written as the pointwise almost everywhere limit of functions of the form $g_1(W_{t_1})g_2(W_{t_2}) \ldots g_k(W_{t_k})$, where $g_1, g_2, \cdots, g_k$ are bounded continuous functions and $t_j \leq t_k$ for $j \leq k, k = 1, 2, \ldots$.*

Definition 9. *Let $\{\mathcal{F}\}_{t\geq 0}$ be an increasing family of σ-algebras of subsets of Ω. A process $g(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}^n$ is called $\mathcal{F}_t$-adapted if for each $t \geq 0$ the function $\omega \mapsto g(t, \omega)$ is $\mathcal{F}_t$-measurable.*

Let us now define very important class of functions.

Definition 10. *A function, say $f(t,\omega) : [0,\infty] \times \Omega \mapsto \mathbb{R}$, is said to be Itō integrable if it satisfies the following conditions:*

(i) $(t,\omega) \mapsto f(t,\omega)$ *is $\mathcal{B} \times \mathcal{F}$-measurable, where $\mathcal{B}$ denotes the Borel σ-algebra on $[0,\infty)$.*

(ii) $f(t,\omega)$ *is $\mathcal{F}_t$-adapted.*

(iii) $E\left[\int_S^T f(t,\omega)^2 dt\right] < \infty$.

Now we can move on to consider the theory for integrals with respect to some stochastic process.

Definition 11. *Let f be Itō integrable. Then the Itō integral of f (from S to T) is defined by*

$$\int_S^T f(t,\omega)dW_t(\omega) = \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)dW_t(\omega) \quad (\textit{limit in } L^2(P))$$

where $\{\phi_n\}$ is a sequence of elementary functions such that

$$E\left[\int_S^T (f(t,\omega) - \phi_n(t,\omega))^2 dt\right] \to 0 \quad \textit{as } n \to \infty.$$

Some key properties of Itō integral are defined in the following propositions.

Proposition 12 (Itō isometry)**.** *Let f be Itō integrable. Then*

$$E\left[\left(\int_S^T f(t,\omega)dW_t\right)^2\right] = E\left[\int_S^T f^2(t,\omega)dt\right].$$

Proposition 13. *Let f, g be Itō integrable and let $0 \le S < U < T$. Then*

(i) $\int_S^T f dW_t = \int_S^U f dW_t + \int_U^T f dW_t$ *almost surely.*

(ii) $\int_S^T (cf+g)dW_t = c \cdot \int_S^T f dW_t + \int_S^T g dW_t$ *almost surely.*

(iii) $E\left[\int_S^T f dW_t\right] = 0$.

(iv) $\int_S^T f dW_t$ *is $\mathcal{F}_t$-measurable.*

By application of the concept of Itō integral, we can look at stochastic processes in the new way.

Definition 14 (1-dimensional Itō process)**.** *Let W_t be 1-dimensional Wiener process on $(\Omega, \mathcal{F}, P)$. A (1-dimensional) Itô process (or stochastic integral) is a stochastic process X_t on $(\Omega, \mathcal{F}, P)$ of the form*

$$X_t = X_0 + \int_0^t u(s,\omega)ds + \int_0^t v(s,\omega)dW_s,$$

where u, v are $\mathcal{F}_t$-adapted, satisfying

$$\Pr\left[\int_0^t v(s,\omega)^2 ds < \infty \quad \textit{for all } t \geq 0\right] = 1.$$

$$\Pr\left[\int_0^t |u(s,\omega)| ds < \infty \quad \textit{for all } t \geq 0\right] = 1.$$

Itō process is often written in shorter differential form $dX_t = udt + vdW_t$. This concept is analogue to ordinary differential equations considering also randomness of a process and is referred to as stochastic differential equation. We can generalize this concept also for higher dimensions.

Definition 15 (Multidimensional Itō process)**.** *Let $W(t,\omega) = (W_1(t,\omega), \cdots, W_m(t,\omega))$ denote m-dimensional Wiener process. If each of the processes $u_i(t,\omega)$ and $v_{ij}(t,\omega)$ satisfies the conditions for 1-dimensional Itô process then we can form the following n Itō processes*

$$\begin{cases} dX_1 = u_1 dt + v_{11}dW_1 + \cdots + & v_{1m}dW_m \\ \vdots \qquad\quad \vdots & \qquad\qquad \vdots \\ dX_n = u_n dt + v_{n1}dW_1 + \cdots + & v_{nm}dW_m \end{cases}$$

Or, in the matrix notation simply $dX(t) = udt + vdW(t)$, where

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{pmatrix}, \quad u = \begin{pmatrix} u_1(t) \\ \vdots \\ u_n(t) \end{pmatrix}, \quad v = \begin{pmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nm} \end{pmatrix}, \quad dW(t) = \begin{pmatrix} dW_1(t) \\ \vdots \\ dW_m(t) \end{pmatrix}.$$

Such a process is called an n-dimensional Itō process.

Basic properties of Itō integral are not very helpful when trying to evaluate its value. Therefore we need to establish stronger tool for computations, something

similar to the role of chain rule in evaluation of Riemann integrals. Itō in [3] came with such a rule, known as Itō formula (or Itō's lemma).

Proposition 16 (The general Itō formula)**.** *Let $dX(t) = udt + vdW_t$ be an n-dimensional Itō process as above. Let $g(t,x) : [0,\infty) \times \mathbb{R}^n \mapsto \mathbb{R}^p$ be C^2 map in x. Then the process $Y(t,\omega) = g(t, X(t))$ is again an Itō process, whose component Y_k is given by*

$$dY_k = \frac{\partial g_k}{\partial t}(t,X)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t,X)dX_i + \frac{1}{2}\frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,X)dX_i \cdot dX_j$$

where we set $dW_i \cdot dW_j = \delta_{ij}dt$ and $dW_i \cdot dt = dt \cdot dW_j = 0$, and δ_{ij} is the Kronecker delta.

At last, we turn to question of existence and uniqueness of the solution of stochastic differential equations.

Proposition 17 (Existence and uniqueness of solutions)**.** *Let $T > 0$ and $b(.,.) : [0,T] \times \mathbb{R}^n \mapsto \mathbb{R}^n, \sigma(.,.) : [0,T] \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times m}$ be measurable functions satisfying $|b(t,x)| + |\sigma(t,x)| \leq C|1 + |x||$; $x \in \mathbb{R}^n, t \in [0,T]$ for some constant C, (where $|\sigma|^2 = \sum |\sigma_{ij}|^2$) and such that $|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|$; $x \in \mathbb{R}^n, t \in [0,T]$ for some constant D. Let Z be a random variable which is independent of the σ- algebra $\mathcal{F}_\infty^{(m)}$ generated by $W_s(.),\ s \geq 0$ and such that $E[|Z|^2] < \infty$. Then the stochastic differential equation $dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t$, $0 \leq t \leq T, X_0 = Z$ has a unique t-continuous solution $X_t(\omega)$ with the property that $X_t(\omega)$ is adapted to the filtration $\mathcal{F}_t^Z$ generated by Z and $W_s(.); s \leq t$ and*

$$E\left[\int_0^T |X_t|^2 dt\right] < \infty.$$

CHAPTER 2

Information theory

2.1 Statistical inference

The aim of this section is to summarize elements of statistical inference relevant for the presented thesis. We refer the reader to the traditional monograph of [6] for the basic theory and article [9] for further topics on regression models.

Definition 18. *Let $f(x,\theta)$ be a joint density function of a random vector X obtained from the experiment. Then the maximum likelihood estimator(MLE) of θ is $\hat{\theta} \in \arg\max_\theta f(X,\theta)$.*

Proposition 19. *If $\hat{\theta}$ is the MLE of θ, then for any function $\kappa(\theta)$, the MLE of $\kappa(\theta)$ is $\kappa(\hat{\theta})$.*

The most used transformation is so-called log-likelihood function and use natural logarithm of a joint density.

Definition 20. *Let $f(x,\theta)$ be a joint density function of a random vector X obtained from an experiment. Then*

$$\mathcal{I}(\theta) = E_X\left[\frac{\partial \ln f(X,\theta)}{\partial \theta}\frac{\partial \ln f(X,\theta)}{\partial \theta^T}\right] \tag{2.1}$$

is called a Fisher information matrix for the unknown parameter θ with respect to observation X.

Proposition 21. *Let us denote the following regularity conditions:*

i) Ω is an open interval (finite, infinite, or semi-infinite).

ii) The distributions F_θ have common support, so that without loss of generality the set $A = \{x : f(x,\theta) > 0\}$ is independent of θ.

iii) For any x in A and θ in Ω, the derivative $f'(x,\theta) = \partial f(x,\theta)/\partial\theta$ and the second derivative $f''(x,\theta) = \partial^2 f(x,\theta)/(\partial\theta\partial\theta^T)$ exist and are finite.

iv) $\frac{\partial}{\partial\theta}\int f(x,\theta)d\mu(x) = \int \frac{\partial}{\partial\theta} f(x,\theta)d\mu(x)$ *and* $\frac{\partial^2}{\partial\theta\partial\theta^T}\int f(x,\theta)d\mu(x) = \int \frac{\partial^2}{\partial\theta\partial\theta^T} f(x,\theta)d\mu(x)$.

Under these conditions, Fisher information matrix can be also evaluated as

$$\mathcal{I}(\theta) = -E_X\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\ln f(X,\theta)\right]. \tag{2.2}$$

Fisher information matrix express the quantity of information that X contains about the parameter θ and has a key value when considering properties of estimation. Variance of the estimation and Fisher information matrix are very closely related as is confirmed by following proposition.

Proposition 22. *Define a parameter column vector θ with probability density function $f(x,\theta)$. Let $T(X)$ be an estimator of any function of parameters and denote its expectation vector $E[T(X)] = \mu(\theta)$. Let us assume that the Fisher information matrix is always defined and the operations of integration with respect to x and differentiation with respect to θ can be interchanged in the expectation of $T(X)$. The Cramér-Rao bound then states that the covariance matrix of $T(X)$ satisfies*

$$Var_\theta[T(X)] \succeq \frac{\partial\mu(\theta)}{\partial\theta^T}[\mathcal{I}(\theta)]^{-1}\frac{\partial\mu^T(\theta)}{\partial\theta}. \tag{2.3}$$

If $T(X)$ is an unbiased estimator of θ (in other words $\mu(\theta) = \theta$), then the Cramér-Rao bound reduces to $Var_\theta[T(X)] \succeq [\mathcal{I}(\theta)]^{-1}$.

The matrix inequality $A \succeq B$ defined on the set of square symmetric matrices denotes Loewner domination and means that the matrix $A - B$ is non-negative definite.

Definition 23. *A random vector x is said to have the multivariate normal distribution with the mean μ and the covariance matrix Σ if its density is in the form*

$$f(x,\mu,\Sigma) = \frac{1}{(2\pi)^{k/2}\det^{1/2}(\Sigma)}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right). \tag{2.4}$$

This definition holds in the non-degenerate case, when covariance matrix Σ is positive definite. Vector x can be also labelled as Gaussian vector. Now let us move to define propositions regarding regression model.

Proposition 24. *We consider a regression model of the form* $y(x_i) = \eta(\theta, x_i) + \varepsilon(x_i)$ *with an unknown vector parameter* θ*. Variance-covariance structure of the observed variables* $y(x_i)$ *is* $Cov(y(x_i), y(x_j)) = \sigma(x_I, x_j; \theta)$*and* $\varepsilon \sim \mathcal{N}(0, \Sigma(\theta))$*. We suppose that the mapping* $\theta \in V \mapsto \eta(\theta) \in \mathbb{R}^N$ *is one-to-one, and the* $N \times N$ *covariance matrix* $C(\theta)$ *with entries* $C(x_i, x_j, \theta)$ *is nonsingular. Suppose also that* $\bar{\theta}$*, the true value of* θ *is from the interior* $int(\Theta)$*. We consider the MLE of the model:* $\hat{\theta} = \arg\max_{\theta\in\Theta} \ln f(y, \theta)$ *where*

$$-\ln f(y,\theta) = \frac{1}{2}\left\{[y-\eta(\theta)]^T\Sigma^{-1}(\theta)[y-\eta(\theta)] + \frac{1}{2}\ln\det[\Sigma(\theta)] + \frac{N}{2}\ln(2\pi)\right\}. \quad (2.5)$$

By taking derivatives, we obtain the Fisher information matrix in the form

$$\mathcal{I}(\theta) = \frac{\partial\eta^T(\theta)}{\partial\theta}\Sigma^{-1}(\theta)\frac{\partial\eta(\theta)}{\partial\theta^T} + \frac{1}{2}tr\left\{\Sigma^{-1}(\theta)\frac{\partial\Sigma(\theta)}{\partial\theta}\Sigma^{-1}(\theta)\frac{\partial\Sigma(\theta)}{\partial\theta^T}\right\}. \quad (2.6)$$

Proposition 25. *Let us consider the model* $y(x_i) = \eta(\nu, x_i) + \varepsilon(x_i)$*. In case of small variances of* $y(x_i)$ *we obtain, that the approximate expression for the MLE is*

$$\hat{\theta} \doteq \bar{\theta} + \mathcal{I}^{-1}(\bar{\theta})\frac{\partial\gamma^T}{\partial\theta}\Big|_{\bar{\theta}}(t-\bar{\mu}).$$

This gives $E_{\bar{\theta}}[\hat{\theta}] \doteq \bar{\theta}$ *and* $Var_{\bar{\theta}}[\hat{\theta}] \doteq \mathcal{I}^{-1}(\bar{\theta})$*.*

2.2 Optimal design of experiments

Here we present some key definitions regarding the theory of experimental design. For more information we refer the reader to the main source of this section, monograph of Pukelsheim [10].

Definition 26. *An experimental design for sample size* n *is given by an* n*-tuple of regression vectors* $\tau_n = (x_1, x_2, \ldots, x_n)^T$*.*

Design for process described by stochastic differential equation consist of n given times in which we observe the process. The vectors that appear in the design τ are called the support of τ.

Definition 27. *Let* S_k^+ *be a set of all* $k\times k$ *non-negative definite matrices. An optimality criterion is a function* $\phi : S_k^+ \mapsto \mathbb{R}$*. It can attain the following properties. (*$\mathcal{I}_1, \mathcal{I}_2 \in S_k^+$*).*

i) Isotonicity relative to the Loewner ordering, $\mathcal{I}_1 \succeq \mathcal{I}_2 \succeq 0 \Rightarrow \phi(\mathcal{I}_1) \geq \phi(\mathcal{I}_2)$*.*

ii) Concavity, $\phi((1-\alpha)\mathcal{I}_1+\alpha\mathcal{I}_2) \geq (1-\alpha)\phi(\mathcal{I}_1)+\alpha\phi(\mathcal{I}_2)$ *for all* $\alpha \in (0,1),\ \ \mathcal{I}_1, \mathcal{I}_2 \succeq 0$.

iii) Positive homogenity, $\phi(\delta\mathcal{I}_1) = \delta\phi(\mathcal{I}_1)$ *for all* $\alpha \in (0,1),\ \ \delta \geq 0,\ \ \mathcal{I}_1 \succeq 0$.

iv) Superadditivity, $\phi(\mathcal{I}_1 + \mathcal{I}_2) \geq \phi(\mathcal{I}_1) + \phi(\mathcal{I}_2)$ *for all* $\mathcal{I}_1, \mathcal{I}_2 \succeq 0$.

v) Upper semicontinuity, the level sets $\{\phi \geq \alpha\} = \{\mathcal{I}_1 :\ \ \phi(\mathcal{I}_1) \geq \alpha\}$ *are closed, for all* $\alpha \in \mathbb{R}$.

We require from the optimality criterion to have these properties so it can capture an idea of whether an information matrix is large or small.

Definition 28. *An information function* ϕ *on* S_k^+ *is a function* $\phi : S_k^+ \mapsto \mathbb{R}$ *that is Loewner isotonic, positively homogenous, superadditive, non-negative, nonconstant and upper semicontinuous.*

Definition 29. *The most prominent information functions are the ones corresponding to the following criteria:*

i) D-criterion, the determinant criterion, $\phi_D(\mathcal{I}) = (det\mathcal{I})^{1/s}$.

ii) A-criterion, the average-variance criterion, $\phi_A(\mathcal{I}) = (\frac{1}{s}tr\{\mathcal{I}^{-1}\})^{-1}$ *if* $\mathcal{I}$ *is positive definite.*

iii) E-criterion, the smallest-eigenvalue criterion, $\phi_E(\mathcal{I}) = \lambda_{\min}(\mathcal{I})$.

Definition 30. *Design* τ_n^* *is said to be optimal, if it satisfies* $\phi(\mathcal{I}(\tau_n^*)) = \sup_{\tau_n \in \mathcal{T}_n} \phi(\mathcal{I}(\tau_n))$.

2.2.1 Ultimate efficiency

In the applications, we usually do not require designs to be optimal, it is enough for them to be in relatively efficient in some way.

Definition 31. *Standard approach to measuring efficiency of a given design* τ *is comparing its information function with the information function of an optimal design.*

$$eff(\tau_n|\phi,\theta) = \frac{\phi[\mathcal{I}(\tau_n,\theta)]}{\sup_{\xi_n \in \mathcal{T}_n} \phi[\mathcal{I}(\xi_n,\theta)]} \tag{2.7}$$

As Pázman [9] suggested, any reference matrix $\mathcal{I}^*$ can be used in the denominator:

$$eff(\tau_n|\phi,\theta) = \frac{\phi[\mathcal{I}(\tau_n,\theta)]}{\phi[\mathcal{I}^*]}, \tag{2.8}$$

where the information matrix $\mathcal{I}^*$ use to be the largest possible in given situation.

Definition 32. *The value* $\lim_{n\to\infty}\mathcal{I}(\tau_n,\theta)$*, with* $\|\tau_n\|\to 0$ *denotes the maximal possible information about the parameter* θ *we can get from the observation of the process at each time from the experimental domain. If this limit exists and is finite then we can measure the design efficiency by*

$$ueff(\tau_n|\phi,\theta)=\frac{\phi[\mathcal{I}(\tau_n,\theta)]}{\lim_{k\to\infty}\phi[\mathcal{I}(\tau_k,\theta)]}. \tag{2.9}$$

The following ratio is known as ultimate efficiency of design(as suggested by Harman [2]).

CHAPTER 3

Multivariate Ornstein-Uhlenbeck process

In this chapter we introduce equations describing wide class of multivariate stochastic processes, known as Ornstein-Uhlenbeck processes. Inspiration for this chapter is work of Lacko [5], currently in press, where this problematic is discussed for univariate processes. As performing computations for multivatiate processes produce higher-dimensional mathematic figures, new problems arise and we need to employ procedures from matrix theory to obtain desired results.

3.1 Formulation of the process

Let us assume a multivariate continuous-time process $\{X(t)\}_{t\geq 0}$, described by linear Itō stochastic differential equation

$$dX(t) = [A_{\upsilon,\beta}(t)X(t) + b_{\upsilon,\beta}(t)]dt + \Sigma_\beta(t)dW_t = f_{\upsilon,\beta}(t, X(t))dt + \Sigma_\beta(t)dW_t,$$
$$X(0) = X_0 \in \mathbb{R} \text{ fixed}, \quad (3.1)$$

where $A_{\upsilon,\beta}(t)$ is a known $n \times n$ matrix which depends on vector parameters υ and β, $b_{\upsilon,\beta}(t)$ is a known $n \times 1$ vector which also depends on υ and β, $\Sigma_\beta(t)$ is known $n \times n$ diffusion matrix dependent only on parameter β and W_t is n-dimensional Wiener process. For the sake of simplicity, we use the notation $\theta = (\upsilon^T, \beta^T)^T$, $\dim(\theta) = m$ and also omit subscriptions from $A(t)$, $b(t)$ and so on. We assume that functions $A(t), b(t)$ and $\Sigma(t)$ and their derivates with respect to θ are integrable with respect to t on the interval $[0, T^*]$ and $\Sigma(t)$ is positive definite. Let us denote the primitive function to $A(t)$ as $\alpha(t)$.

The process governed by (3.1) is often referred to as multivariate Ornstein-Uhlenbeck process.

In the present setup we can observe the process $\{X(t)\}_{t\geq 0}$ at n strictly increasing design times $\tau = (t_1, t_2, \ldots, t_n)^T$ from the experimental domain $\mathcal{D} = [T_*, T^*]$. We denote the set of all feasible n-point designs by $\mathcal{T}_{n,\mathcal{D}} = \{\tau \in \mathbb{R}^n;\ T_* \leq t_1 < t_2 < \cdots < t_n \leq T^*\}$, and $\overline{\mathcal{T}}_{n,\mathcal{D}} = \{\tau \in \mathbb{R}^n;\ T_* \leq t_1 \leq t_2 \leq \cdots < t_n \leq T^*\}$ be its closure. We further define $\|\tau\| = \max_{2\leq i\leq n}(t_i - t_{i-1})$ the norm of the partition generated by the sampling design τ.

3.2 Properties of the multivariate Ornstein-Uhlenbeck process

By using the transformation $Y(t) = e^{-\alpha(t)}X(t) = g(t,x)$, Itō's lemma yields

$$dY(t) = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial X}dX + 0$$

$$dY(t) = e^{-\alpha(t)}b(t)dt + e^{-\alpha(t)}\Sigma(t)dW_t$$

$$Y(t) - Y(t_0) = e^{-\alpha(t)}X(t) - e^{-\alpha(t_0)}X(t_0) = \int_{t_0}^{t} e^{-\alpha(s)}b(s)ds + \int_{t_0}^{t} e^{-\alpha(s)}\Sigma(s)dW_s$$

$$X(t)|X(t_0) = e^{\alpha(t)}e^{-\alpha(t_0)}X(t_0) + \int_{t_0}^{t} e^{\alpha(t)}e^{-\alpha(s)}b(s)ds + \int_{t_0}^{t} e^{\alpha(t)}e^{-\alpha(s)}\Sigma(s)dW_s.$$

From the basic properties of Itō integral and Itō's isometry we obtain the expression for mean and covariance matrix of $X(t)$.

Lemma 33. *It holds true that*

$$E[X(t)|X(t_0)] = E[t|t_0] = e^{\alpha(t)}e^{-\alpha(t_0)}X(t_0) + \int_{t_0}^{t} e^{\alpha(t)}e^{-\alpha(s)}b(s)ds \tag{3.2}$$

and

$$Var[X(t)|X(t_0)] = Var[t|t_0] = \int_{t_0}^{t} e^{\alpha(t)}e^{-\alpha(s)}\Sigma(s)\Sigma^T(s)(e^{\alpha(t)}e^{-\alpha(s)})^T ds. \tag{3.3}$$

We use simplified version of the notation for E and Var introduced in this lemma through the whole chapter.

For further dicsussion on experimental design, it is desirable to write the information matrix about the unknown parameter θ in more practical way.

Lemma 34. *Let $\{X(t)\}_{t\geq 0}$ with $X(0)$ fixed be a multivariate θ-parametrised continuous-time Gaussian Markov process. Then for any $\tau \in \mathcal{T}_{n,\mathcal{D}}$, the Fisher information matrix for $X(\tau)$ takes the form*

$$\mathcal{I}(\tau,\theta) = \mathcal{I}_{X(t_1)|X(0)}(\theta) + \sum_{i=2}^{n} E_{X(t_{i-1})}\left[\mathcal{I}_{X(t_i)|X(t_{i-1})}(\theta)\right], \tag{3.4}$$

where $\mathcal{I}_{X(t_i)|X(t_{i-1})}$ denotes the Fisher information matrix for X_{t_i} conditioned on the value $X(t_{i-1})$ and $E_{X(t_{i-1})}[.]$ is the expectation with respect to $X(t_{i-1})$.

Proof. We refer reader to work of Lacko [5], as the proof is straightforward generalization of the univariate case. □

This lemma provides very powerful tool for computing Fisher information matrix as in the multivariate case if we wanted to compute an information matrix for $X(\tau)$ as a whole we would require to know the covariancne matrix for $X(\tau)$, which could be very challenging as the covariance of one observation is already a matrix and therefore it would be a matrix of matrices. In this way, we simplified our task to compute Fisher information matrix for each observation, conditioned in the previous observation.

Now let us take a closer look on $\mathcal{I}_{X(t_{i+1})|X(t_i)}(\theta)$.

Definition 35. *Fisher information matrix for observation in time t_{i+1} conditioned in time t_i takes the form*

$$\mathcal{I}(t_{i+1}|t_i,\theta) = E_{X(t_i)}\left[\left[\frac{\partial E[t_{i+1}|t_i]}{\partial \theta^T}\right]^T Var^{-1}[t_{i+1}|t_i]\frac{\partial E[t_{i+1}|t_i]}{\partial \theta^T}\right] +$$

$$+\frac{1}{2}tr\left\{Var^{-1}[t_{i+1}|t_i]\frac{\partial Var[t_{i+1}|t_i]}{\partial \theta}Var^{-1}[t_{i+1}|t_i]\frac{\partial Var[t_{i+1}|t_i]}{\partial \theta^T}\right\}$$

where $E_{X(t)}$ denotes expectation with respect to observation in time t.

We need to take a closer look at the expression inside the trace, as it is the product of $n \times n$ matrix, $n \times m \times n$ $3-$dimensional tensor, $n \times n$ matrix again and finally another $3-$dimensional tensor$(m \times n \times n)$. Result is a $4-$dimensional tensor with dimensions $n \times m \times m \times n$. Since the result has to be $m \times m$ matrix, we can define this trace as

$$\frac{1}{2}tr\{.\} = \frac{1}{2}\sum_{k=1}^{n}(Var^{-1}[t_{i+1}|t_i])_{k.}\frac{\partial Var[t_{i+1}|t_i]}{\partial \theta}Var^{-1}[t_{i+1}|t_i]\frac{\partial (Var[t_{i+1}|t_i])_{.k}}{\partial \theta^T}. \tag{3.5}$$

3.3 Existence of optimal design

At this point we turn our attention to the existence of feasible optimal design for the parameter θ. From Lemma 33 is easy to see that $E[X(t)|X(t)] = X(t)$ almost surely, thus we do not gain additional information about the parameter from replication of the observation. Therefore if we perform the replicated observation in the different time and leave original non-replicated observations, from the form of Fisher information matrix introduced in Lemma 34 is clear that we increase the amount of information we get from an experiment. It follows that for any design from the boundary $\overline{\mathcal{T}}_{n,\mathcal{D}} \setminus \mathcal{T}_{n,\mathcal{D}}$ there exist a desing from $\mathcal{T}_{n,\mathcal{D}}$ which dominates it. Therefore we can reduce the set of competing designs to $\mathcal{T}_{n,\mathcal{D}}$. The conventional way to decide, whether an optimal design exists is that it achieves maximum on the set of competing designs. But as the set of designs $\mathcal{T}_{n,\mathcal{D}}$ is not compact, information function $\phi[\mathcal{I}(X(\tau), \theta)]$ does not have to reach its maximum on this set. Problems may appear in the case, when Fisher information matrix is not continuous on the boundary $\overline{\mathcal{T}}_{n,\mathcal{D}} \setminus \mathcal{T}_{n,\mathcal{D}}$ as Lacko suggested in [5]. We use the definition of existence of an optimal n-point sampling design in the strong sense, which is satisfied, if the following statement holds: for any boundary design $\tau_0 \in \overline{\mathcal{T}}_{n,\mathcal{D}} \setminus \mathcal{T}_{n,\mathcal{D}}$, there exists a design $\tau \in \mathcal{T}_{n,\mathcal{D}}$, such that for any sequence of designs $\{\tau^{(k)}\}$ on $\mathcal{T}_{n,\mathcal{D}}$ with $\lim_{k\to\infty} \tau^{(k)} = \tau_0$, the Fisher information matrix for τ Loewner dominates the matrix $\lim_{k\to\infty} \mathcal{I}(\tau^{(k)}, \theta)$. Thus the existence of optimal design also does not depend on the choice of the information function $\phi[\mathcal{I}]$. For further examination of the limit properties of Fisher information matrix as the observation times go near each other we need to express $\lim_{\Delta\to 0} \mathcal{I}_{X(t+\Delta)|X(t)}(\theta)$. For the asymptotic estimation of this matrix as $\Delta \to 0$ we use Taylor expansion up to order one given as following: $f(t + \Delta) = f(t) + f'(t)\Delta + o(\Delta)$. Let us start with individual components of $\mathcal{I}_{X(t+\Delta)|X(t)}(\theta)$.

For the better understanding of the following computations, let us establish known Leibniz's rule for differentiation under the integral sign. For reference regarding definitions from calculus we refer the reader to [4].

Proposition 36. *Let $f(x, \theta)$ be a function such that derivative $f_\theta(x, \theta)$ exists and is continuous. Then,*

$$\frac{d}{d\theta}\left(\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx\right) = \int_{a(\theta)}^{b(\theta)} f_\theta(x,\theta)dx + f(b(\theta),\theta)b'(\theta) - f(a(\theta),\theta)a'(\theta).$$

Using Taylor expansion as $\Delta \to 0$ for $Var[t+\Delta|t]$ with the help of Proposition 36 we obtain

$$\begin{aligned} Var[t+\Delta|t] &\doteq 0 + e^{\alpha(t+\Delta)}e^{-\alpha(t+\Delta)}\Sigma(t+\Delta)\Sigma^T(t+\Delta)\left(e^{\alpha(t+\Delta)}e^{-\alpha(t+\Delta)}\right)^T\Big|_{\Delta=0}\Delta + \\ &+o(\Delta) = \Sigma(t)\Sigma^T(t)\Delta + o(\Delta). \end{aligned}$$

It seems to be more difficult with the other element of Fisher information matrix

$$\frac{\partial}{\partial\theta^T}E[t+\Delta|t] = \frac{\partial}{\partial\theta^T}\left(e^{\alpha(t+\Delta)}e^{-\alpha(t)}X(t) + \int_t^{t+\Delta} e^{\alpha(t+\Delta)}e^{-\alpha(s)}b(s)ds\right)$$

If we add a special zero $\left(\pm e^{\alpha(t+\Delta)}e^{-\alpha(t)}E[X(t)]\right)$ and realise that $X(t)$ and $E[X(t)]$ do not depend directly on θ, we can rewrite our formula as following:

$$\frac{\partial}{\partial\theta^T}E[t+\Delta|t] = \frac{\partial}{\partial\theta^T}\left[e^{\alpha(t+\Delta)}e^{-\alpha(t)}\right](X(t)-E[X(t)]) + \frac{\partial}{\partial\theta^T}\left[e^{\alpha(t+\Delta)}e^{-\alpha(t)}\right]E[X(t)] +$$

$$+\frac{\partial}{\partial\theta^T}\int_t^{t+\Delta} e^{\alpha(t+\Delta)}e^{-\alpha(s)}b(s)ds.$$

Let us now define some important regularity conditions, known also as Schwarz theorem.

Proposition 37 ([4]). *Suppose that f is a function of two variables such that $\frac{\partial^2 f(X,Y)}{\partial X^T \partial Y^T}$ and $\frac{\partial^2 f(X,Y)}{\partial Y^T \partial X^T}$ both exist and are continuous at some point (X_0, Y_0). Then*

$$\frac{\partial^2 f(X,Y)}{\partial X^T \partial Y^T}\Big|_{X_0,Y_0} = \frac{\partial^2 f(X,Y)}{\partial Y^T \partial X^T}\Big|_{X_0,Y_0}.$$

Under regularity conditions mentioned above we can change the order of partial derivation and therefore first compute derivation with respect to t coming from Taylor expansion and then resolve partial derivation with respect to θ. First we figure out Taylor series of $e^{\alpha(t+\Delta)}e^{-\alpha(t)}$ as $\Delta \to 0$.

$$\begin{aligned} e^{\alpha(t+\Delta)}e^{-\alpha(t)} &\doteq I + e^{\alpha(t+\Delta)}A(t+\Delta)e^{-\alpha(t)}\Big|_{\Delta=0}\Delta + o(\Delta) \\ &= I + e^{\alpha(t)}A(t)e^{-\alpha(t)}\Delta + o(\Delta) \\ &= I + A(t)\Delta + o(\Delta). \end{aligned}$$

As we work everywhere with matrices, which in general do not commute $AB \neq BA$, we have to show the validity of the operation in the last step. First is the well known identity $e^A \cdot e^{-A} = I$. For the sake of the other identity let us formulate the following lemma.

Lemma 38. *If A is a regular matrix, such that its spectral decomposition exists, the following equation holds: $A \cdot e^A = e^A \cdot A$.*

Proof. Eigendecomposition of A can be given as $A = V\Lambda V^{-1}$, where $\Lambda = diag(\lambda_i)$ and V is an orthogonal matrix of corresponding eigenvectors. Identity $A^k = V\Lambda^k V^{-1}$ together with the fact, that matrix exponential can be evaluated as the sum of increasing power functions implies that $e^A = e^{V\Lambda V^{-1}} = Ve^{\Lambda}V^{-1} = Vdiag(e^{\lambda_i})V^{-1}$. Therefore $Ae^A = V\Lambda V^{-1}Ve^{\Lambda}V^{-1} = V\Lambda e^{\Lambda}V^{-1} = Ve^{\Lambda}\Lambda V^{-1} = e^A A$. (Since Λ and e^{Λ} are diagonal matrices, they can be switched) □

In the next part we calculate Taylor series for $\int_t^{t+\Delta} e^{\alpha(t+\Delta)}e^{-\alpha(s)}b(s)ds$ as $\Delta \to 0$. Using Proposition 36 we get

$$\begin{aligned}\int_t^{t+\Delta} e^{\alpha(t+\Delta)}e^{-\alpha(s)}b(s)ds &\doteq 0 + e^{\alpha(t+\Delta)}e^{-\alpha(t+\Delta)}b(t+\Delta)\Big|_{\Delta=0}\Delta + o(\Delta)\\ &= b(t)\Delta + o(\Delta).\end{aligned}$$

If we sum it up together with the fact that $\frac{\partial I}{\partial \theta^T} = 0$ we came to the following result:

$$\frac{\partial}{\partial \theta^T}E[t+\Delta|t] = \left[\left(\frac{\partial A(t)}{\partial \theta^T}\right)E[X(t)] + \frac{\partial b(t)}{\partial \theta^T}\right]\Delta +$$

$$+\left(\frac{\partial A(t)}{\partial \theta^T}\right)(X(T) - E[X(t)])\Delta + o(\Delta)$$

Now we are ready to construct asymptotic estimate of Fisher information matrix. We will refer to the summands of the formula for $\frac{\partial}{\partial \theta^T}E[t+\Delta|t]$ as Summand 1 and Summand 2. Now let us concentrate on the first part of $\mathcal{I}(t+\Delta|t,\theta)$ which is now in the form

$$E_{X(t)}\left[\left[\frac{\partial E[t+\Delta|t]}{\partial \theta^T}\right]^T Var^{-1}[t+\Delta|t]\frac{\partial E[t+\Delta|t]}{\partial \theta^T}\right] =$$

$$\begin{aligned}= \left[\left[\left(\frac{\partial A(t)}{\partial \theta^T}\right)E[X(t)] + \frac{\partial b(t)}{\partial \theta^T}\right]\Delta + o(\Delta)\right]^T \cdot \left[\Sigma(t)\Sigma^T(t)\Delta + o(\Delta)\right]^{-1} \cdot\\ \cdot \left[\left[\left(\frac{\partial A(t)}{\partial \theta^T}\right)E[X(t)] + \frac{\partial b(t)}{\partial \theta^T}\right]\Delta + o(\Delta)\right] +\\ +E_{X(t)}\left[(X(t) - E[X(t)])^T\left(\frac{\partial A(t)}{\partial \theta^T}\right)^T \cdot \left[\Sigma(t)\Sigma^T(t)\Delta + o(\Delta)\right]^{-1} \cdot\right.\\ \left.\cdot\left(\frac{\partial A(t)}{\partial \theta^T}\right)(X(T) - E[X(t)])\Delta^2 + o(\Delta^2)\right].\end{aligned}$$

In the substitution into the formula we used already obtained Taylor expansions together with the fact that $E_{X(t)}[X(t)] = E[X(t)]$. It implies that products of type 'Summand1' $\cdot Var^{-1}[t+\Delta|t]\cdot$ 'Summand2' equal to zero and therefore can be omitted from the formula.

Considering asymptotic properties up to order 1, we may regard $\frac{o(\Delta)}{\Delta}$ as 0. We also recall the formula for the drift of the process $A(t)X(t) + b(t) = f(X(t))$. Seeing, that our result is a matrix of dimension $m \times m$. We can write down its ij-th component as :

$$\frac{\partial f(X(t))^T}{\partial \theta_i}\Big|_{E[X(t)]} \left[\Sigma(t)\Sigma^T(t)\right]^{-1} \frac{\partial f(X(t))}{\partial \theta_j}\Big|_{E[X(t)]}\Delta +$$

$$+E_{X(t)}\Bigg[(X(t) - E[X(t)])^T \left(\frac{\partial A(t)}{\partial \theta_i}\right)^T \left[\Sigma(t)\Sigma^T(t)\right]^{-1} \cdot$$

$$\cdot \left(\frac{\partial A(t)}{\partial \theta_j}\right)(X(T) - E[X(t)])\Bigg]\Delta + o(\Delta).$$

We would like to find a way to simplify the following expression to more favourable form. Inside the $E_{X(t)}\Big[.\Big]$ is a scalar, which means that $E_{X(t)}\Big[.\Big] = E_{X(t)}\Big[tr\{.\}\Big]$. Using the known identity of a trace, $tr\{AB\} = tr\{BA\}$, we obtain

$$E_{X(t)}\Big[.\Big] = E_{X(t)}\Bigg[tr\Bigg\{\left(\frac{\partial}{\partial \theta_i}A(t)\right)^T \left[\Sigma(t)\Sigma^T(t)\right]^{-1} \left(\frac{\partial}{\partial \theta_j}A(t)\right) \cdot$$

$$\cdot (X(T) - E[X(t)])\,(X(t) - E[X(t)])^T\Bigg\}\Bigg] =$$

$$= tr\Bigg\{\left(\frac{\partial A(t)}{\partial \theta_i}\right)^T \left[\Sigma(t)\Sigma^T(t)\right]^{-1} \left(\frac{\partial A(t)}{\partial \theta_j}\right) \cdot Var[X(t)]\Bigg\}.$$

In the last step we use basic definition for the variation $\mathrm{Var}[X] = E[(X - E(X))(X - E(X))^T]$.

Now we can proceed to the second part of of the conditioned information matrix, which equals to

$$\frac{1}{2}tr\left\{Var^{-1}[t_{i+\Delta}|t_i]\frac{\partial Var[t_{i+\Delta}|t_i]}{\partial \theta}Var^{-1}[t_{i+\Delta}|t_i]\frac{\partial Var[t_{i+\Delta}|t_i]}{\partial \theta^T}\right\}.$$

From the Taylor expansion we get it simplified as following

$$\frac{1}{2}tr\Bigg\{\left[\Sigma(t)\Sigma^T(t)\cdot\Delta + o(\Delta)\right]^{-1} \cdot \frac{\partial \Sigma(t)\Sigma^T(t)\cdot\Delta + o(\Delta)}{\partial \theta}.$$

$$\cdot \left[\Sigma(t)\Sigma^T(t)\cdot\Delta + o(\Delta)\right]^{-1} \cdot \frac{\partial \Sigma(t)\Sigma^T(t)\cdot\Delta + o(\Delta)}{\partial\theta^T}\Bigg\}.$$

We can observe that inside the trace is expression in the form

$$((.)\cdot\Delta + o(\Delta))^{-1}\cdot((.)\cdot\Delta + o(\Delta))\cdot((.)\cdot\Delta + o(\Delta))^{-1}\cdot((.)\cdot\Delta + o(\Delta)),$$

which contains members independent to Δ and therefore this part tends to some non-zero matrix. Let us now summarize the result of our computations in the following lemma.

Lemma 39.

$$\lim_{\Delta\to 0}\left(E_{X(t)}\left[\mathcal{I}_{t+\Delta|t}(\theta)\right] - \mathcal{I}_1(t,\theta)\Delta - \mathcal{I}_2(t,\theta)\right) = \mathbf{0}_{m\times m}, \tag{3.6}$$

where

$$\mathcal{I}_1(t,\theta) = \frac{\partial f^T(X(t))}{\partial\theta}\Big|_{E[X(t)]}\left[\Sigma(t)\Sigma^T(t)\right]^{-1}\frac{\partial f(X(t))}{\partial\theta^T}\Big|_{E[X(t)]} + \\ + tr\left\{\frac{\partial A^T(t)}{\partial\theta}\left[\Sigma(t)\Sigma^T(t)\right]^{-1}\frac{\partial A(t)}{\partial\theta^T}\cdot Var[X(t)]\right\}$$

and

$$\mathcal{I}_2(t,\theta) = \frac{1}{2}tr\left\{\left[\Sigma(t)\Sigma^T(t)\right]^{-1}\frac{\partial\Sigma(t)\Sigma^T(t)}{\partial\theta}\left[\Sigma(t)\Sigma^T(t)\right]^{-1}\cdot\frac{\partial\Sigma(t)\Sigma^T(t)}{\partial\theta^T}\right\}.$$

Inside the traces in the formulas are again 4-dimensional $n\times m\times m\times n$ tensors, therefore sumation has to be carry out throughout the dimensions with n members.

As $\mathcal{I}_1$ is coefficient associated with Δ, this part tends to zero as observations from τ converges to τ_0 and therefore this part is continuous on the boundary $\overline{\mathcal{T}}_{n,\mathcal{D}}\setminus\mathcal{T}_{n,\mathcal{D}}$.

But since $\mathcal{I}_2$ does not come with Δ in the formula, it can cause discontinuity on the boundary $\overline{\mathcal{T}}_{n,\mathcal{D}}\setminus\mathcal{T}_{n,\mathcal{D}}$. But recalling the fact that Σ does not depend on the parameter υ, $\mathcal{I}_2$ does not influence components of Fisher information matrix depending on any parameter from υ. Hence discontinuity can be caused only by the block $\mathcal{I}_{\beta\beta}$ of Fisher information matrix. We are now ready to state one of the biggest achieved results in the thesis.

Theorem 40. *If β is known parameter of equation* (3.1)*, describing multivariate Ornstein-Uhlenbeck process, then an optimal sampling design for υ exists in the strong sense.*

Proof. If β is not an unknown parameter of our stochastic differential equation, potentional discontinuity is avoided and thus $\lim_{\tau\to\tau_0}\mathcal{I}(\tau,\theta) = \mathcal{I}(\tau_0,\theta)$. Together

with the fact, that for any design with replications τ_0 exists feasible design from its neighbouhood, such that $\mathcal{I}(\tau, \theta) \succeq \mathcal{I}(\tau_0, \theta)$ it yields existence of the achieved maximum on the set $\mathcal{T}_{n,\mathcal{D}}$ and thus also existence of optimal sampling design in the strong sense. □

For better grasp of the problems associated with the parameter β present in the volatility, we recall the fractal property of Wiener process, which implies that it keeps exactly the same properties, regardless of the "zoom" we use. Therefore if we move the observations closer to each other, the amount of information about the parameters linked to Wiener process and thus present in the volatility does not decrease and does not tend to zero as observations converge to each other. This fractal property is also illustrated on Figure 3.1, zoom of the process increases from upper left subplot to lower right subplot.

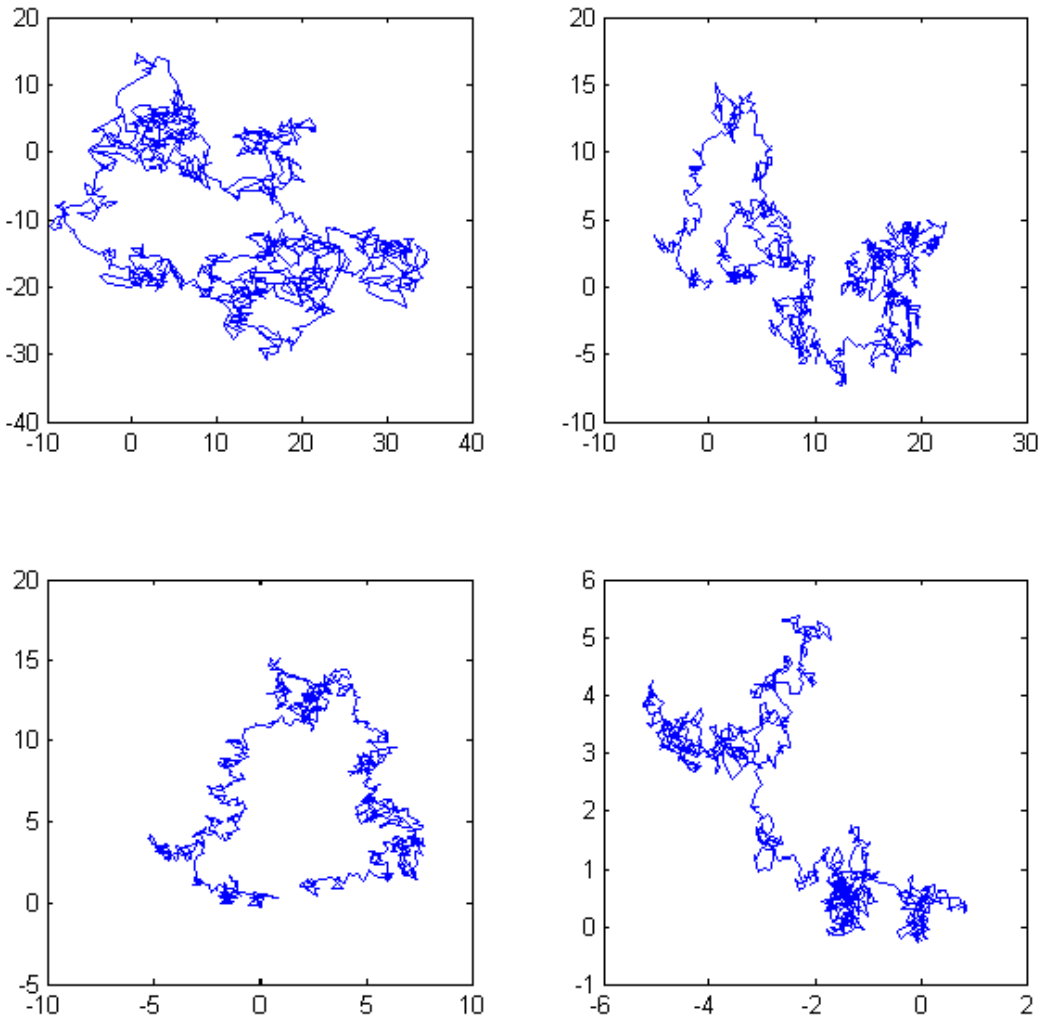

Figure 3.1: *Fractal property of* $2-$*dimensional Wiener process*

3.4 Ultimate efficiency

As we mentioned earlier, as a reference matrix to measure quality of designs by so-called ultimate efficiency is Fisher information matrix obtained by the observation of the trajectory at all times from the experimental domain. This matrix $\mathcal{I}_\infty(\theta) = \lim_{n\to\infty} \mathcal{I}(\tau^{(n)}, \theta)$ as $\|\tau^{(n)}\| \to 0$ is often referred to as asymptotic Fisher information matrix. Corresponding design $\tau^{(n)}$ is said to cover the whole experimental domain $\mathcal{D}$. Reasonable values of ultimate efficiency can be only obtained in the case when $\lim_{n\to\infty} \phi[\mathcal{I}(\tau^{(n)}, \theta)]$ attains finite value. We already know all important facts to formulate theorem crucial for computing asymptotic information matrix, also being one of the main results of the thesis.

Theorem 41. *Let $\{X(t)\}_{t\geq 0}$ be a process governed by multivariate stochastic differential equation* (3.1)*, and let $\tau^{(n)}$ covers the whole experimental domain $\mathcal{D}$. Then*

$$\lim_{n\to\infty} \left(\mathcal{I}(\tau^{(n)}, \theta) - \mathcal{I}_\infty(\theta) - \sum_{i=2}^{n} \mathcal{I}_2(t_i, \theta) \right) = \mathbf{0}_{m\times m}, \tag{3.7}$$

where

$$\mathcal{I}_\infty(\theta) = \frac{\partial E^T[T_*]}{\partial \theta} Var^{-1}[T_*] \frac{\partial E[T_*]}{\partial \theta^T} +$$

$$+\frac{1}{2} tr \left\{ Var^{-1}[T_*] \frac{\partial Var[T_*]}{\partial \theta} Var^{-1}[T_*] \frac{\partial Var[T_*]}{\partial \theta^T} \right\} + \int_{T_*}^{T^*} \mathcal{I}_1(t, \theta) dt$$

and $\mathcal{I}_1(t, \theta)$ and $\mathcal{I}_2(t, \theta)$ are defined in Lemma 39*.*

Proof. Lemma 34 and Lemma 39 together implies the statement of the theorem. □

As the Fisher information matrix for the whole design is computed through the sums for individual observations, let us introduce the following proposition for better understanding of the summation of multiple matrices.

Proposition 42 ([7])**.** *Let A and B be Hermitian matrices in M_n with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$, respectively. Then, if $d_1 \geq d_2 \geq \cdots \geq d_n$ denote the diagonal entries of $A + B$, we have that the vector of the diagonal elements of $A + B$ is majorized by the sum of the vectors of the eigenvalues of A and B. This means that*

$$d_1 + \cdots + d_k \leq \lambda_1 + \mu_1 + \cdots + \lambda_k + \mu_k, \qquad k = 1, \cdots, n-1$$

and

$$d_1 + \cdots + d_n = \lambda_1 + \mu_1 + \cdots + \lambda_n + \mu_n.$$

This proposition also includes easier situation, where $B = 0$ and thus vector of the diagonal elements of any Hermitian matrix is majorized by the vector of its eigenvalues. If we want to apply this lemma on $\mathcal{I}_2$, first we have to ensure its symetry. So we demand the following equation to hold: $\{\mathcal{I}_2\}_{ij} = \{\mathcal{I}_2\}_{ji}$. By recalling (3.5) and the identity $tr\{AB\} = tr\{AB\}$ we get (in the simplified form) $\{\mathcal{I}_2\}_{ij} = \frac{\partial \mathrm{Var}}{\partial \theta_i}_{k.} \mathrm{Var}^{-1} \frac{\partial \mathrm{Var}}{\partial \theta_j} \mathrm{Var}^{-1}_{.k}$ and $\{\mathcal{I}_2\}_{ji} = \mathrm{Var}^{-1}_{k.} \frac{\partial \mathrm{Var}}{\partial \theta_j} \mathrm{Var}^{-1} \frac{\partial \mathrm{Var}}{\partial \theta_i}_{.k}$. As transposition of scalar is the same scalar value and thanks to symetry of matrices Var^{-1} and $\frac{\partial \mathrm{Var}}{\partial \theta_j}$ we can assume $\mathcal{I}_2$ to be symetric. From the form of $\mathcal{I}_2$ we can say, that it is also positive definite. It implies that some eigenvalues and hence also information about $\{\lim_{n\to\infty} \mathcal{I}(\tau^{(n)}, \theta)\}_{\beta\beta}$ tends to infinity.

Recalling the substitution $\theta = (\upsilon^T, \beta^T)^T$ we can write information matrix in the block form $\mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_{\upsilon\upsilon} & \mathcal{I}_{\upsilon\beta} \\ \mathcal{I}_{\beta\upsilon} & \mathcal{I}_{\beta\beta} \end{pmatrix}$. Fisher information matrix corresponding to subparameter υ can be evaluated as Schur complement for the block $\mathcal{I}_{\beta\beta}$.

$$\mathcal{I}_\upsilon(\tau, \theta) = \mathcal{I}_{\upsilon\upsilon} - \mathcal{I}_{\upsilon\beta} \left(\mathcal{I}_{\beta\beta}\right)^- \mathcal{I}_{\beta\upsilon}.$$

Analogically

$$\mathcal{I}_\beta(\tau, \theta) = \mathcal{I}_{\beta\beta} - \mathcal{I}_{\beta\upsilon} \left(\mathcal{I}_{\upsilon\upsilon}\right)^- \mathcal{I}_{\upsilon\beta}.$$

In both cases $\mathcal{I}^-$ denotes arbitrary pseudo-inverse of $\mathcal{I}$.

As achieved information about β is not bounded, it can be estimated consistently and therefore the concept of ultimate efficiency is not applicable here.

Now let us focus on subparameter υ. Information about $\{\lim_{n\to\infty} \mathcal{I}(\tau^{(n)}, \theta)\}_{\beta\beta}$ tend to infinity and thus $\mathcal{I}_{\upsilon\beta} \left(\mathcal{I}_{\beta\beta}\right)^- \mathcal{I}_{\beta\upsilon} \to \mathbf{0}_{m\times m}$ and $\mathcal{I}_\upsilon(\tau^{(n)}, \upsilon) \to \{\mathcal{I}_\infty(\theta)\}_{\upsilon\upsilon}$, which implies expected property $ueff(\tau^{(n)}|\phi, \upsilon) \to 1$ as $n \to \infty$. To sum it up, with $\mathcal{I}_\infty(\theta)$ defined in Theorem 41, we can compute ultimate efficiency of design τ for estimation of subvector υ of the parameter vector θ as

$$ueff(\tau|\phi, \upsilon) = \frac{\phi[\mathcal{I}_\upsilon(\tau, \theta)]}{\phi[\{\mathcal{I}_\infty(\theta)\}_{\upsilon\upsilon}]}.$$

Conclusion

In the first two chapters we summarized the theory of stochastic calculus and probability and information theory used later on. In the main chapter of the thesis we formulated stochastic model for multivariate Ornstein-Uhlenbeck processes, discussed some of their properties and using unconventional methods we evaluated Fisher information matrix. We studied limit case, when two observations approach each other and with the help of our results we conditioned the existence of optimal sampling design by knowledge of the parameter present in the volatility of the stochastic process, in our notation presented as β. We also stated that the concept of ultimate efficiency obtains reasonable outcome when we consider estimation of parameters that do not contain β. Obtained results can be further applicated, for example in the field of multi-compartment models, describing wide class of processes in pharmacokinetics or biomedicine.

Bibliography

[1] Gompertz, B.: *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*, Philos. Trans. R. Soc. Lond. 115(1825), 513-585.

[2] Harman, R., Štulajter, F.: *Optimal sampling designs for the Brownian Motion with a quadratic drift*, Journal of Statistical Planning and Inference 141(2011), 2480-2488.

[3] Itô, K.: *On a formula concerning stochastic differentials*, Nagoya Mathematical Journal 3(1951), 55-65.

[4] James, R.C.: *Advanced Calculus*, Wadsworth, 1966.

[5] Lacko, V.: *Ultimate efficiency of experimental designs for Ornstein-Uhlenbeck type processes*, in press: http://dx.doi.org/10.1016/j.jspi.2014.02.003i .

[6] Lehmann, E.L., Casella, G.: *Theory of Point Estimation*, Springer, 1998.

[7] Miranda, H.: *Diagonals and eigenvalues of Hermitian matrices. Extreme cases*, Scientific Electronic Library Online, 2003.

[8] Øksendal, B.: *Stochastic Differential Equations*, Springer, 2000.

[9] Pázman, A.: *Criteria for optimal design of small-sample experiments with correlated observations*, Kybernetika 43(2007), 453-462.

[10] Pukelsheim, F.: *Optimal Design of Experiments*, Wiley, 1993.

[11] Uhlenbeck, G.E., Ornstein, L. S.: *On the Theory of the Brownian Motion*, Physical Review 36(1930), 823-841.

[12] Sacks, J., Ylvisaker, D.: *Designs for regression problems with correlated errors*, Ann. Math. Stat. 39(1966), 66-89.