Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Optimum design in nonlinear models

Dissertation thesis

2018 **Mgr. Katarína Sternműllerová**

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Optimum design in nonlinear models

Dissertation thesis

Study programme:	Applied mathematics
Field of study:	1114 Applied mathematics
Department:	Department of Applied Mathematics and Statistics
Supervisor:	prof. RNDr. Andrej Pázman, DrSc.

Bratislava 2018

Mgr. Katarína Sternműllerová
(nee Burclová)

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

33148456

THESIS ASSIGNMENT

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

Title: Optimum design in nonlinear models

Annotation: Our aim is to transform some optimality criteria, so to be directly applicable in the optimization via linear programming methods, and so that the class of optimality criteria used in nonlinear models will be meaningfully extended. Namely, we would like to use the I-divergence in the models described by the exponential families of distributions and also apply some knowledges from the theory of risk.

Aim: To generalize some results of the paper, and Chapters 7 and 8 of the monograph included in the list of references.

Literature: Pázman, A. and Pronzato, L. (2014). Optimum design accounting for the global nonlinear behavior of the model. The Annals of Statistics, 42(4):1426–1451.

Pronzato, L. and Pázman, A. (2013). Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties. Lecture Notes in Statistics, Vol. 212. Springer, New York, Heidelberg.

Tutor:	prof. RNDr. Andrej Pázman, DrSc.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, DrSc.
Assigned:	10.02.2014

Approved: 20.02.2014 prof. RNDr. Daniel Ševčovič, DrSc.
Guarantor of Study Programme

.....
Student

.....
Tutor

Abstract

STERNMŰLLEROVÁ, Katarína. *Optimum design in nonlinear models.* Comenius University in Bratislava. Faculty of Mathematics, Physics and Informatics; Department of Applied Mathematics and Statistics. Supervisor: prof. RNDr. Andrej Pázman, DrSc. Bratislava 2018, 132 pages.

Keywords: optimality criteria, the cutting plane method, Conditional Value at Risk, Kullback-Leibler I-divergence, exponential family of distributions.

Abstract:

The thesis deals with some new approaches of optimal experimental design in nonlinear models. Following the paper Pázman and Pronzato (2014) and the monograph Pronzato and Pázman (2013), we construct new forms of optimality criteria, we investigate their mathematical properties, and we demonstrate the possibility of obtaining optimal experimental designs using the methods of linear programming.

In the thesis we extend the criteria which are considered in Pázman and Pronzato (2014) and are related to the stability of the least square estimate in a nonlinear regression model. Namely, applying the I-divergence, we can at the design stage of the experiment reach the improvement of the stability of the maximum likelihood estimate in a generalized regression model based on the exponential family of distributions. In addition, we formulate some other optimality criteria which follow similar purposes but are closely related to different well-known optimality criteria not considered in Pázman and Pronzato (2014).

Further, we elaborate the issues of the criterion based on the Conditional Value at Risk, which was used in optimal experimental design by Valenzuela et al. (2015) for the first time. We analyse this criterion from the point of view of the optimal experimental design and we use linear programming to calculate optimal designs.

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Optimalizácia experimentov v nelineárnych modeloch

Dizertačná práca

Študijný program:	Aplikovaná matematika
Študijný odbor:	1114 Aplikovaná matematika
Pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	prof. RNDr. Andrej Pázman, DrSc.

Bratislava 2018

Mgr. Katarína Sternműllerová
(rod. Burclová)

33148456

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

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

Názov: Optimum design in nonlinear models
Optimalizácia experimentov v nelineárnych modeloch

Anotácia: Našim cieľom je pretransformovať niektoré kritériá optimality tak, aby boli dobre využiteľné pri optimalizácii metódami lineárneho programovania a tak, aby bola zmysluplne rozšírená trieda kritérií optimality v nelineárnych modeloch. Menovite chceme využiť I-divergenciu v modeloch opisovaných exponenciálnymi triedami rozdelení pravdepodobnosti ako aj aplikovať niektoré poznatky známe z teórie rizika.

Cieľ: Zovšeobecniť niektoré výsledky článku a Kapitol 7 a 8 z monografie uvedených v zozname literatúry.

Literatúra: Pázman, A. and Pronzato, L. (2014). Optimum design accounting for the global nonlinear behavior of the model. The Annals of Statistics, 42(4):1426–1451.

Pronzato, L. and Pázman, A. (2013). Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties. Lecture Notes in Statistics, Vol. 212. Springer, New York, Heidelberg.

Školiteľ: prof. RNDr. Andrej Pázman, DrSc.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, DrSc.

Dátum zadania: 10.02.2014

Dátum schválenia: 20.02.2014 prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
školiteľ

Abstrakt

STERNMŰLLEROVÁ, Katarína. *Optimalizácia experimentov v nelineárnych modeloch.* Univerzita Komenského v Bratislave. Fakulta matematiky, fyziky a informatiky; Katedra aplikovanej matematiky a štatistiky. Vedúci práce: prof. RNDr. Andrej Pázman, DrSc. Bratislava 2018, 132 strán.

Kľúčové slová: kritériá optimality, metóda cutting plane, podmienená hodnota v riziku, Kullbackova-Leiblerova I-divergencia, exponenciálna trieda rozdelení.

Abstrakt:

Predkladaná záverečná práca sa venuje niektorým novým prístupom navrhovania experimentov v nelineárnych modeloch. Vychádzajúc z článku Pázman a Pronzato (2014) a z monografie Pronzato a Pázman (2013) formulujeme nové tvary kritérií optimality, analyzujeme ich matematické vlastnosti a prezentujeme možnosť získania optimálnych návrhov experimentov aplikovaním metód lineárneho programovania.

V práci rozšírime kritériá optimality, ktoré boli uvažované v Pázman a Pronzato (2014) a súvisia so stabilitou odhadu metódou najmenších štvorcov v nelineárnom regresnom modeli. Konkrétne s využitím I-divergencie vieme v štádiu navrhovania experimentu dosiahnuť zlepšenie stability odhadu metódou maximálnej vierohodnosti v zovšeobecnenom regresnom modeli založenom na exponenciálnej triede rozdelení. Navyše, sformulujeme ďalšie kritériá optimality, ktoré plnia podobný účel, ale sú úzko späté s inými známymi kritériami optimality neuvažovanými v článku Pázman a Pronzato (2014).

Ďalej rozpracujeme problematiku kritéria založeného na podmienenej hodnote v riziku, ktoré bolo po prvý raz aplikované v oblasti navrhovania experimentov v článku Valenzuela et al. (2015). Analyzujeme toto kritérium z hľadiska optimálneho navrhovania experimentov a opäť použijeme lineárne programovanie na výpočet optimálnych návrhov.

Acknowledgements

I would like to thank my supervisor, prof. RNDr. Andrej Pázman, DrSc., for creating the conditions for the study and research, lending me the relevant literature, for his time provided during our discussions and for his important notes which caused significant improvements of this thesis.
My thanks go also to Luc Pronzato, Directeur de Recherche at CNRS, France, for his helpful advises and ideas.
I am very thankful to my friends and colleagues for their support during my studies. Especially, I would like to express my gratitude to my husband and parents, who have always been here for me, for their care and persistent encouragements. Thank you.

Katarína Sternműllerová

Contents

List of abbreviations xii

List of symbols xiii

Preface 1

1 Preliminaries 4
- 1.1 The basic matrix algebra results . . . 4
- 1.2 Linear and nonlinear regression models . . . 5
- 1.3 Experimental design in nonlinear models . . . 5
- 1.4 The cutting plane method . . . 8
  - 1.4.1 A subgradient of a concave function . . . 9

2 Experimental design via linear programming 11
- 2.1 Reformulation of criteria of E-, **c**-, and G-optimality . . . 12
- 2.2 Reformulation of criteria of D-, A-, and E_k-optimality . . . 12
- 2.3 Algorithm and example . . . 16
  - 2.3.1 Computation of criterion robust design . . . 19
  - 2.3.2 Computation of D-optimal design conditioned by a prescribed level of A-optimality . . . 21
- 2.4 Reformulation of AVE criteria . . . 21
- 2.5 Bibliographic remarks on LP methods in experimental design . . . 22
- 2.6 Conclusions to this chapter . . . 23

3 Experimental design based on some ideas from risk theory 24
- 3.1 Basic definitions . . . 25
- 3.2 CVaR-criterion for experimental design purposes . . . 27
  - 3.2.1 Concavity of the CVaR-criterion . . . 27
  - 3.2.2 Conditions for finiteness of the CVaR-criterion . . . 29
  - 3.2.3 Continuity of the function $w_\alpha(\xi, c)$ in the formulation of the CVaR-criterion in (3.9) . . . 30

3.2.4 Optimal c in the definition of CVaR-criterion Φ_α in (3.7) 30
3.2.5 The dependence of the CVaR-criterion Φ_α on the probability level α 31
3.3 Two easily interpreted criteria as lower and upper bound of the CVaR-criterion 32
3.3.1 Formulation of some optimality criteria based on the risk theory 33
3.3.2 The comparison of the criteria based on the risk theory 35
3.3.3 The CVaR-criterion in the case of continuous random variable $\phi(\xi, \boldsymbol{\theta})$ 39
3.4 Interpretation of the CVaR-criterion Φ_α in the context of experimental design 42
3.4.1 Relation to the local, maximin, and AVE-criterion 43
3.4.2 Relation to the quantile criterion 44
3.4.3 Further properties of the CVaR-criterion 45
3.5 Calculation of CVaR-optimal designs 48
3.5.1 Subgradient for the function $w_\alpha(\xi, c)$ from (3.9) 49
3.5.2 The algorithm 51
3.6 Examples 52
3.7 Conclusions to this chapter 64

4 Extended optimality criteria for avoiding false estimates in generalized regression models 66
4.1 Introduction 67
4.1.1 Matrix Norms and Pseudonorms 67
4.1.2 Exponential families of distributions 69
4.1.3 Generalized regression models based on exponential families . . 72
4.2 Experimental design in generalized regression models 76
4.2.1 Designing experiments in generalized regression models using the asymptotic properties of maximum likelihood estimate 76
4.3 The problem of stability and identifiability of maximum likelihood estimate. The extended optimality criteria. 78
4.3.1 The general definition of extended optimality criteria 80
4.4 The extension of pseudonorm optimality criteria 81
4.4.1 The case when $k = 1$ 87
4.4.2 The case when $k = m$ and the distance $\rho(\cdot; \cdot, \cdots, \cdot)$ is given through Schatten p norm 90
4.4.3 The case when $k < m$ and the distance $\rho(\cdot; \cdot, \cdots, \cdot)$ is given through a pseudonorm 93

4.5 Alternative extensions of **c**- and G-optimality criteria in generalized regression models . . . 94
4.6 Properties of extended optimality criteria defined in (4.21) . . . 95
4.6.1 Parameter K . . . 96
4.7 Algorithm and examples . . . 96
4.8 Conclusions of new results in this chapter . . . 102

Conclusion **104**

Resumé **106**
Úvod . . . 106
Navrhovanie experimentov v nelineárnom regresnom modeli . . . 106
Navrhovanie experimentov v zovšeobecnenom regresnom modeli založenom na exponenciálnej triede rozdelení . . . 107
Ciele dizertačnej práce . . . 108
Výsledky práce . . . 109
Navrhovanie experimentov pomocou lineárneho programovania . . . 109
Navrhovanie experimentov pomocou kritérií inšpirovaných teóriou rizika 111
Formulácia rozšírených kritériií optimality za účelom obmedzenia mylných odhadov v zovšeobecnených regresných modeloch . . . 114

Appendix **117**
A Reformulation of L-optimality criterion . . . 117
B Some examples of distributions from the exponential family . . . 119
B.1 Binomial distribution $Bin(n,p)$ . . . 119
B.2 Poisson distribution $Po(\lambda)$ . . . 120
B.3 Geometric distribution $Ge(p)$ . . . 120
B.4 Negative binomial distribution $NBin(r,p)$ . . . 121
B.5 Exponential distribution $Exp(\lambda)$ . . . 121
B.6 Weibull distribution $W(\lambda,k)$ . . . 122
B.7 Pareto distribution $Pa(\alpha,\sigma)$ . . . 122
B.8 Normal distribution $\mathcal{N}\left(\nu,\sigma^2\right)$ with σ^2 given . . . 123
B.9 Normal distribution $\mathcal{N}\left(\nu,\sigma^2\right)$ . . . 123
B.10 Gamma distribution $G(k,\alpha)$ . . . 124
B.11 Beta distribution $Be(a,b)$ . . . 125
B.12 Multinomial distribution $Mult(n,p_1,\ldots,p_{l-1})$ . . . 126

Bibliography **127**

List of abbreviations

AVE	average
cdf	cumulative distribution function
CVaR	Conditional Value at Risk
LP	linear programming
pdf	probability density function
pmf	probability mass function
VaR	Value at Risk

List of symbols

$\mathbb{N}$	the set of natural numbers $\{1, 2, \ldots\}$
$\mathbb{R}^p$	the set of all p-dimensional real vectors
$\mathbb{R}$	$\mathbb{R}^1$
$\mathbb{R}^{m\times k}$	the set of all $m \times k$-dimensional real matrices
$a, b, c, \lambda, \ldots$	scalars
$\mathbf{a}, \mathbf{u}, \mathbf{v}, \ldots$	vectors
$A, B, M, Q, U, \ldots$	matrices
$\mathrm{diag}\{a_1, \ldots, a_m\}$	the diagonal matrix with $a_1, \ldots, a_m$ on the main diagonal
I	$\mathrm{diag}\{1, \ldots, 1\}$
$A_{\cdot j}$	the j-th column of the matrix A
$\{A\}_{ij}$	the element in the i-th row and j-th column of the matrix A
$\mathrm{tr}(A)$	the trace of the matrix A
$\det(A)$	the determinant of the matrix A
$\lambda_{\min}(A)$	the smallest eigenvalue of the matrix A
$\lambda_1(M) \leq \ldots \leq \lambda_m(M)$	the ordered eigenvalues of the symmetric matrix $M \in \mathbb{R}^{m\times m}$
A^{-1}	the inverse of the matrix A
A^-	a g-inverse of the matrix A
Pr	probability
$\mathcal{N}(a, \sigma^2)$	the one-dimensional normal distribution with mean a and variance σ^2
$\lvert a\rvert$	the absolute value of a
$\lVert\mathbf{a}\rVert, \lVert A\rVert$	a vector norm of the vector $\mathbf{a}$, a matrix norm of the matrix A
$\lVert\cdot\rVert^D$	the dual norm of a norm $\lVert\cdot\rVert$
$\lVert\mathbf{a}\rVert_{\ell(p)}$	the $\ell(p)$ norm of the vector $\mathbf{a}$
$\lVert A\rVert_{\mathrm{S}(p)}$	the Schatten p norm of the square matrix A

$\|\|\|\mathbf{a}\|\|\|_P, \|\|\|A\|\|\|_P$	a vector pseudonorm of the vector $\mathbf{a}$, a matrix pseudonorm of the matrix A
$\mathcal{C}(A) \equiv \left\{A\mathbf{u} : \mathbf{u} \in \mathbb{R}^k\right\}$	the column space of the matrix $A \in \mathbb{R}^{m\times k}$
$\mathbf{0}$	the null vector or null matrix
$Y, \mathbf{Y}$	a random variable, a random vector
F_Y	cdf of a random variable Y
P_Y	pmf of a random variable Y
f_Y	pdf of a random variable Y
$q_Y(\alpha)$	the α-left-quantile of Y
$Rq_Y(\alpha)$	the α-right-quantile of Y
$\mathcal{Y} \subseteq \mathbb{R}^l$	the sample space
$\mathbf{y}$	an element of $\mathcal{Y}$
τ	a measure on $\mathcal{Y}$
$\boldsymbol{\vartheta}$	a parametrization in the exponential family
$t(\mathbf{y}) = (t_1(\mathbf{y}), \ldots, t_r(\mathbf{y}))^\top$	the sufficient statistic in the exponential family
$\gamma(\boldsymbol{\vartheta}) = (\gamma_1(\boldsymbol{\vartheta}), \ldots, \gamma_r(\boldsymbol{\vartheta}))^\top$	the canonical function in the exponential family
$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)^\top$	the canonical parameter in the exponential family
Γ	the canonical space
$\Gamma_{\max}$	the maximal canonical space
$\mathrm{int}(\Gamma)$	the interior of Γ
$f(\mathbf{y}, \boldsymbol{\gamma})$	the pdf or pmf of an exponential family
$E(\cdot)$	the expected value
$E(Y\|Y < c)$	the expected value of Y under the condition that $Y < c$
$E_{\boldsymbol{\gamma}}(\cdot)$	the expected value for given $\boldsymbol{\gamma}$
$\mathrm{Var}(\cdot)$	the variance or covariance matrix
$\mathrm{Var}_{\boldsymbol{\gamma}}(\cdot)$	the variance or covariance matrix for given $\boldsymbol{\gamma}$
$\overline{\mu}(\boldsymbol{\gamma})$	$E_{\boldsymbol{\gamma}}[t(\mathbf{Y})]$
$\overline{\Sigma}(\boldsymbol{\gamma})$	$\mathrm{Var}_{\boldsymbol{\gamma}}[t(\mathbf{Y})]$
$\mathcal{X} \subseteq \mathbb{R}^d$	the design space
$\mathrm{card}(\mathcal{X})$	the cardinality of $\mathcal{X}$
$\mathbf{x}$	the design point, an element of $\mathcal{X}$
$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^\top$	the regression parameter
$\bar{\boldsymbol{\theta}}$	the true value of the parameter $\boldsymbol{\theta}$
$\boldsymbol{\theta}^0$	the nominal parameter value
$\hat{\boldsymbol{\theta}}_N$	the least squares or maximum likelihood

	estimate for $\boldsymbol{\theta}$ based on N measurements
$L(\boldsymbol{\theta})$	the likelihood function
$\Theta \subseteq \mathbb{R}^m$	the regression parametric space
Θ^2, Θ^k	$\Theta \times \Theta$, $\Theta \times \Theta \times \ldots \times \Theta$
π	the prior distribution of $\boldsymbol{\theta}$ on Θ
$\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$	the vector of regression functions in a linear regression model
$\eta(\mathbf{x}, \boldsymbol{\theta}) \in \mathbb{R}$	the expected value of an observation in a nonlinear regression model
ε	the random error in a linear or nonlinear regression model
$f(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$	the density of an measurement in the generalized regression model
$g(\mathbf{x}, \boldsymbol{\theta}) = (g_1(\mathbf{x}, \boldsymbol{\theta}), \ldots, g_r(\mathbf{x}, \boldsymbol{\theta}))^\top$	the regression function in a generalized regression model
$\mu(\mathbf{x}, \boldsymbol{\theta}) = (\mu_1(\mathbf{x}, \boldsymbol{\theta}), \ldots, \mu_r(\mathbf{x}, \boldsymbol{\theta}))^\top$	$\overline{\mu}[g(\mathbf{x}, \boldsymbol{\beta})]$
$\Sigma(\mathbf{x}, \boldsymbol{\beta})$	$\overline{\Sigma}[g(\mathbf{x}, \boldsymbol{\beta})]$
$\mathscr{E}$	the expectation surface
$\mathscr{K}$	the canonical surface
N	the number of measurements in the experiment
$X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$	the exact experimental design
$\xi, \zeta, \nu, \ldots$	the (aproximative) experimental design
$\delta_{\mathbf{x}}$	the Dirac measure concentrated at $\mathbf{x}$
Ξ	the set of all experimental designs defined on $\mathcal{X}$
$M(\mathbf{x}, \boldsymbol{\theta})$	the elementary (Fisher) information matrix
$M(\xi, \boldsymbol{\theta})$	the (Fisher) information matrix associated with the design ξ
$M_X(\boldsymbol{\theta})$	the (Fisher) information matrix associated with the exact design X
$I_{\mathbf{x}}(\boldsymbol{\theta}^0, \boldsymbol{\theta})$	the elementary I-divergence for given $\mathbf{x}$
$I_X(\boldsymbol{\theta}^0, \boldsymbol{\theta})$	the I-divergence associated with the exact design X
$\bar{\phi}(\cdot)$	the optimality criterion defined on the set of information matrices
$\phi(\cdot, \cdot)$	the optimality criterion defined on $\Xi \times \Theta$
$\phi^{ext}(\cdot, \cdot)$	the extended optimality criterion defined on $\Xi \times \Theta$

$\Phi_{\mathrm{loc}}, \Phi_{\min}, \Phi_{\mathrm{AVE}}$	the local, maximin and AVE optimality criteria
Φ_α^Q	the quantile criterion for given $\alpha \in [0,1]$
Φ_α	the CVaR criterion for given $\alpha \in (0,1]$
$\nabla\varphi(\widetilde{\mathbf{z}})$	a subgradient a of a concave function φ at $\widetilde{\mathbf{z}}$
$\nabla_\xi \phi(\widetilde{\xi}, \boldsymbol{\theta})$	a subgradient of $\phi(\cdot, \boldsymbol{\theta})$ at $\widetilde{\xi}$
$\mathcal{F}_\Phi(\xi,\nu), \mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu)$	a directional derivative of the criterion $\xi \mapsto \Phi(\xi)$ and $\xi \mapsto \phi(\xi, \boldsymbol{\theta})$, respectively, at ξ in the direction ν
$O(\cdot)$	the big O notation

More symbols are defined in the text.

Preface

In the thesis we describe some possibilities of designing experiments in nonlinear models, where the information matrix and the criterion function depend on the unknown parameter that is aimed to be estimated. From the statistical point of view, the experiment is optimal if it brings most information on the unknown parameter, often under some financial or material restrictions.

The monographs Pronzato and Pázman (2013) or Fedorov and Leonov (2014) are devoted to optimal experimental design in nonlinear regression models and provide a deep analysis of many design methods. In the thesis we try to extend some of these known methods and we also pay attention on one approach, which has not been sufficiently studied in the area of optimal experimental design. A common element which is related to all the methods considered in the thesis is linear programming—largely used here to obtain optimal designs (the cutting plane method, cf. Kelley, 1960).

Structure and aim of the thesis

Chapter 1 Preliminaries contains some basic and very general results from the matrix theory, regression models, optimal experimental design and describes the above-mentioned cutting plane method.

The remaining chapters can be considered separately. The aim of each chapter is indicated at its beginning. Every chapter also contains some introductory notes (known from the literature) and the conclusion, which summarizes the achieved results.

Chapter 2 Experimental design via linear programming presents the results of Burclová and Pázman (2016a). The main goals:

- to reformulate the criteria of D-, A-, E-, and E_k-optimality into the form which allows to use linear programming for optimal experimental design (Theorems 2.1, 2.2, and 2.3),

- to demonstrate that the iterative LP method can be used to solve corresponding design problems numerically (Sect. 2.3), and
- to use these reformulations in more complicated problems like the calculation of criterion robust design (Sect. 2.3.1) or optimal design under some subsidiary linear constraints (Sect. 2.3.2).

Chapter 3 Experimental design based on some ideas from risk theory studies the possibilities of applying results from the risk theory into the area of optimal experimental design as suggested in the paper Valenzuela et al. (2015). The main goals:

- to define the criterion based on Conditional Value at Risk (CVaR) as a concave function of the design ξ and consider also discrete prior distributions (Sect. 3.2),
- to provide an analysis of the CVaR criterion based on the theory of risk in order to obtain its meaningful interpretation (Sect. 3.3),
- to study its relation to AVE, maximin, local (Theorem 3.19) and quantile criteria (Sect. 3.4.2), which is also done in the examples (mainly Example 3.25),
- to derive the directional derivative (Lemma 3.20) and to prove the equivalence theorem for the CVaR criterion (Theorem 3.21), and
- to demonstrate the possibility of obtaining CVaR-optimal designs via linear programming (Sect. 3.5).

Chapter 4 Extended optimality criteria for avoiding false estimates in generalized regression models extends the results of Pázman and Pronzato (2014) to more general regression models defined by exponential families and partially is based on Burclová and Pázman (2016b). The main tool used here is an adequate adaptation of the I-divergence. **Appendix A** involves one further theoretical result used in this chapter and **Appendix B** presents some distributions from the exponential family. The main goals of the chapter:

- to redefine the extended criteria of E-, $\mathbf{c}$-, and G-optimality from Pázman and Pronzato (2014) so that they are applicable in generalized regression models based on the exponential family (Sect. 4.5 and Eq. (4.27)),
- to elaborate the necessary mathematical tools to obtain extended criteria also for the MV-, L-, and some L_p-optimality criteria, not considered earlier (Sects. 4.4.1–4.4.3),

- to prove that the considered criteria are really extensions of the classical optimality criteria (Theorems 4.16, 4.17, and 4.18), and
- to apply the results of previous chapters to optimal experimental design in generalized regression models based on the exponential family of distributions (Sect. 4.2.1).

Chapter 1

Preliminaries

1.1 The basic matrix algebra results

The eigenvalues $\lambda_1(M) \leq \lambda_2(M) \leq \ldots \leq \lambda_m(M)$ of any symmetric matrix $M \in \mathbb{R}^{m \times m}$ are real. Every symmetric matrix M can be written as $M = Q\Lambda Q^\top$, where the matrix Λ is diagonal with the eigenvalues $\lambda_1(M), \ldots, \lambda_m(M)$ on the main diagonal and Q is an orthogonal matrix, the columns of which contain the orthonormally chosen eigenvectors of M corresponding to $\lambda_1(M), \ldots, \lambda_m(M)$ (see e.g. Harville, 2008, Sects. 21.4–21.5 or Gentle, 2007, Chap. 3.8).

For any p one defines the p-th power of a symmetric positive semidefinite matrix M as $M^p = Q\Lambda^p Q^\top$, since the eigenvalues on the diagonal of Λ are always nonnegative (see e.g. Gentle, 2007, Chap. 3.8).

We formulate the Cauchy-Schwarz inequality for matrices.

Theorem 1.1. *(Harville, 2008, Theorem 6.3.1) Let A and B be matrices in $\mathbb{R}^{m \times k}$. Then one has*

$$\mathrm{tr}^2\left(A^\top B\right) \leq \mathrm{tr}\left(A^\top A\right) \mathrm{tr}\left(B^\top B\right).$$

As a consequence of Theorem 1.1 one obtains for a symmetric positive definite matrix $M = M^{1/2} M^{1/2} \in \mathbb{R}^{m \times m}$ (after substitution $A \leftarrow M^{1/2} A$, $B \leftarrow M^{-1/2} B$) that for any $A \neq \mathbf{0}$, B

$$\frac{\mathrm{tr}^2\left(A^\top B\right)}{\mathrm{tr}\left(A^\top M A\right)} \leq \mathrm{tr}\left(B^\top M^{-1} B\right) \tag{1.1}$$

with the equality sign for $A = M^{-1} B$, hence

$$\max_{A \in \mathbb{R}^{m \times k} : A \neq \mathbf{0}} \frac{\mathrm{tr}^2\left(A^\top B\right)}{\mathrm{tr}\left(A^\top M A\right)} = \mathrm{tr}\left(B^\top M^{-1} B\right). \tag{1.2}$$

We mention one more important relation from the matrix theory (see e.g. Harville, 2008, Theorem 21.5.6)

$$\min_{\mathbf{u}:\|\mathbf{u}\|_{\ell(2)}=1} \mathbf{u}^\top M \mathbf{u} = \lambda_{\min}(M), \tag{1.3}$$

where $\lambda_{\min}(M)$ denotes the minimal eigenvalue of a symmetric positive semidefinite matrix M and $\|\cdot\|_{\ell(2)}$ is the Euclidean norm.

1.2 Linear and nonlinear regression models

Consider the linear regression model

$$y(\mathbf{x}) = \mathbf{f}^\top(\mathbf{x})\boldsymbol{\theta} + \varepsilon, \tag{1.4}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^\top \in \mathbb{R}^m$ is a vector of regression parameters, $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$ is a known continuous vector function and the regressor $\mathbf{x}$ is from the design space $\mathcal{X} \subset \mathbb{R}^d$. The design space $\mathcal{X}$ is throughout the thesis assumed to be compact unless otherwise stated. The random errors ε are supposed to have zero mean and, generally unknown, constant variance σ^2.

Denote by X an exact experimental design consisting of N points form $\mathcal{X}$, i.e. $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. If we perform independent measurements in points from X with results $y(\mathbf{x}_1), \ldots, y(\mathbf{x}_N)$, then the covariance matrix of least squares estimate for $\boldsymbol{\theta}$ equals $\sigma^2 M_X^{-1}$ if the information matrix $M_X = \sum_{i=1}^N \mathbf{f}(\mathbf{x}_i)\mathbf{f}^\top(\mathbf{x}_i)$ associated with the exact design X is nonsingular.

When the expectation of $y(\mathbf{x})$ is not linear in $\boldsymbol{\theta}$, then $y(\mathbf{x})$ may satisfy a nonlinear regression model of the form

$$y(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon, \tag{1.5}$$

where η is a known mapping $\eta : \mathcal{X} \times \Theta \mapsto \mathbb{R}$ continuous on Θ, a compact subset of $\mathbb{R}^m$. If $\eta(\mathbf{x}, \boldsymbol{\theta})$ is moreover differentiable on $\operatorname{int}(\Theta)$, we define for $\boldsymbol{\theta}^0 \in \operatorname{int}(\Theta)$ the elementary information matrix as a function of $\mathbf{x} \in \mathcal{X}$ and $\boldsymbol{\theta}^0$

$$M\left(\mathbf{x}, \boldsymbol{\theta}^0\right) = \left.\frac{\partial \eta(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \left.\frac{\partial \eta(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}. \tag{1.6}$$

Throughout the thesis we assume that any two observations $y(\mathbf{x})$, $y'(\mathbf{x})$ from (1.4) or (1.5) corresponding to different trials are independent.

More general regression models are considered in Chap. 4.

1.3 Experimental design in nonlinear models

The issues of optimal experimental design are in detail studied e.g. in Pronzato and Pázman (2013) or Fedorov and Leonov (2014). The aim of this section is only to

explain the basic terms and notations.

The approximative experimental design (hereinafter referred to as „design“) is any probability measure ξ defined on $\mathcal{X}$, and the set of all such measures is denoted by Ξ. The information matrix associated with a discrete design ξ is defined as

$$M\left(\xi, \boldsymbol{\theta}^0\right) = \sum_{\mathbf{x} \in \mathcal{X}: \xi(\mathbf{x}) > 0} M\left(\mathbf{x}, \boldsymbol{\theta}^0\right) \xi\left(\mathbf{x}\right), \tag{1.7}$$

where $M\left(\mathbf{x}, \boldsymbol{\theta}^0\right)$ is the elementary information matrix defined in (1.6). For brevity, we sometimes use the notation $\begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots \\ \xi(\mathbf{x}_1) & \xi(\mathbf{x}_2) & \dots \end{Bmatrix}$ for a discrete design ξ.

Throughout this thesis, if no confusion occurs, we will use the same notation for the information matrix $M(\xi, \boldsymbol{\theta})$, as a function of a design $\xi \in \Xi$, and for the elementary information matrix $M(\mathbf{x}, \boldsymbol{\theta})$, as a function of a design point $\mathbf{x} \in \mathcal{X}$, although the latter one should be expressed as $M(\delta_{\mathbf{x}}, \boldsymbol{\theta})$, where $\delta_{\mathbf{x}} \in \Xi$ is a Dirac measure concentrated at $\mathbf{x}$.

If $\xi' \in \Xi$ is not a discrete measure, one may write in (1.7) an integral over $\mathcal{X}$ instead of the sum to obtain $M\left(\xi', \boldsymbol{\theta}^0\right)$. However, as a consequence of Caratheodory's theorem, see e.g. Pronzato and Pázman (2013), Sect. 5.2.3, $M\left(\xi', \boldsymbol{\theta}^0\right) = M\left(\xi, \boldsymbol{\theta}^0\right)$ for some discrete design ξ.

In Chap. 3 of Pronzato and Pázman (2013) is proved that when $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ is a random sample from distribution given by ξ, then, under some assumptions including a nonsingularity of $M\left(\xi, \bar{\boldsymbol{\theta}}\right)$, the covariance matrix of the least square estimate of $\boldsymbol{\theta}$ based on independent measurements in $\mathbf{x}_1, \dots, \mathbf{x}_N$ is asymptotically, for $N \to \infty$, proportional to $\frac{1}{N}\left[M\left(\xi, \bar{\boldsymbol{\theta}}\right)\right]^{-1}$, where $\bar{\boldsymbol{\theta}} \in \operatorname{int}(\Theta)$ is the true parameter value. One sees that the asymptotic covariance matrix depends on the unknown true parameter value $\bar{\boldsymbol{\theta}}$. A more direct and simple proof is possible if $\mathcal{X}$ is a finite set. In that case it is sufficient that the relative frequency of any point $\mathbf{x}$ within X is converging to $\xi(\mathbf{x})$, see e.g. Pázman and Lacko (2012), Sect. 3.7.

Usually, the main idea of statistical optimization of experiments lies in maximizing (with respect to ξ) a chosen optimality criterion, which measures the largeness of the information matrix associated with given design ξ (and hence minimizes the covariance of the least squares estimate). An optimality criterion in a nonlinear model is a real function ϕ of the design ξ and of the parameter $\boldsymbol{\theta}$ given as $\phi(\xi, \boldsymbol{\theta}) = \bar{\phi}\left[M\left(\xi, \boldsymbol{\theta}\right)\right]$ for some function $\bar{\phi}$ defined on the set of all information matrices. For purposes of optimal experimental design we will assume that $M(\mathbf{x}, \boldsymbol{\theta})$ is continuous on $\mathcal{X}$ (this is always true when $\mathcal{X}$ is finite) for given $\boldsymbol{\theta}$.

Below we define some optimality criteria which are for any $\boldsymbol{\theta} \in \Theta$

- concave, i.e. $\phi[\lambda\xi_1+(1-\lambda)\xi_2,\boldsymbol{\theta}] \geq \lambda\phi(\xi_1,\boldsymbol{\theta})+(1-\lambda)\phi(\xi_2,\boldsymbol{\theta})$ for any $\xi_1,\xi_2 \in \Xi$ and for any $\lambda \in (0,1)$, and
- positively homogeneous, i.e. $\phi(a\xi,\boldsymbol{\theta}) = a\phi(\xi,\boldsymbol{\theta})$ for any $\xi \in \Xi$ and for any $a > 0$.

Definition 1.2. Let $\xi \in \Xi$, $\boldsymbol{\theta} \in \Theta$. In a nonlinear regression model we define the

criterion of D-optimality:

$$\phi_D(\xi,\boldsymbol{\theta}) = \{\det[M(\xi,\boldsymbol{\theta})]\}^{\frac{1}{m}},$$

criterion of A-optimality:

$$\phi_A(\xi,\boldsymbol{\theta}) = \begin{cases} 1/\mathrm{tr}\left\{[M(\xi,\boldsymbol{\theta})]^{-1}\right\} & \text{if } M(\xi,\boldsymbol{\theta}) \text{ is nonsingular,} \\ 0 & \text{otherwise,} \end{cases}$$

criterion of E_k-optimality:

$$\phi_{E_k}(\xi,\boldsymbol{\theta}) = \sum_{i=1}^{k} \lambda_i[M(\xi,\boldsymbol{\theta})],$$

where $\lambda_1[M(\xi,\boldsymbol{\theta})] \leq \lambda_2[M(\xi,\boldsymbol{\theta})] \leq \ldots \leq \lambda_m[M(\xi,\boldsymbol{\theta})]$ are the ordered eigenvalues of matrix $M(\xi,\boldsymbol{\theta})$ respecting their multiplicity,

criterion of E-optimality:

$$\phi_E(\xi,\boldsymbol{\theta}) = \lambda_1[M(\xi,\boldsymbol{\theta})] = \lambda_{\min}[M(\xi,\boldsymbol{\theta})],$$

criterion of c-optimality:

$$\phi_{\mathbf{c}}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\mathbf{c}^\top[M(\xi,\boldsymbol{\theta})]^{-}\mathbf{c}} & \text{if } \mathbf{c} \in \mathcal{C}[M(\xi,\boldsymbol{\theta})], \\ 0 & \text{otherwise,} \end{cases}$$

where $[M(\xi,\boldsymbol{\theta})]^{-}$ is an arbitrarily chosen g-inverse of matrix $[M(\xi,\boldsymbol{\theta})]$,

criterion of G-optimality:

$$\phi_G(\xi,\boldsymbol{\theta}) = \begin{cases} \min_{\mathbf{x}\in\mathcal{X}} \frac{1}{\mathbf{f}^\top(\mathbf{x})[M(\xi,\boldsymbol{\theta})]^{-1}\mathbf{f}(\mathbf{x})} & \text{if } M(\xi,\boldsymbol{\theta}) \text{ is nonsingular,} \\ 0 & \text{otherwise.} \end{cases}$$

For the interpretation of criteria from Def. 1.2 in linear models see e.g. Pázman (1986) or Pukelsheim (1993) and in nonlinear models see Chap. 5 of Pronzato and Pázman (2013). The criteria of E_k optimality were defined and applied in Harman (2004).

Since at the design stage of the experiment the true parameter value $\bar{\boldsymbol{\theta}}$ is unknown, we are not able to maximize $\phi\left(\xi, \bar{\boldsymbol{\theta}}\right)$. Hence there is a need to avoid a dependence of $\phi(\xi, \boldsymbol{\theta})$ on the unknown parameter value $\boldsymbol{\theta}$ and the usual way is to consider one of the below formulated criteria.

Definition 1.3. Suppose that $\phi(\xi, \boldsymbol{\theta})$ is a real function defined on $\Xi \times \Theta$, e.g. one of the criteria introduced in Def. 1.2. Then we define the

local optimality criterion

$$\Phi_{\mathrm{loc}}(\xi) = \phi\left(\xi, \boldsymbol{\theta}^0\right), \tag{1.8}$$

where $\boldsymbol{\theta}^0$ is the nominal parameter value which is a priori assumed to be in the neighbourhood of $\bar{\boldsymbol{\theta}}$,

maximin optimality criterion

$$\Phi_{\min}(\xi) = \min_{\boldsymbol{\theta} \in \Theta} \phi(\xi, \boldsymbol{\theta}), \tag{1.9}$$

average (AVE) optimality criterion

$$\Phi_{\mathrm{AVE}}(\xi) = \int_{\Theta} \phi(\xi, \boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}), \tag{1.10}$$

where $\pi(\cdot)$ is some prior probability distribution on the parametric space Θ.

The concavity and positive homogeneity of $\phi(\xi, \boldsymbol{\theta})$ in ξ then imply the concavity and positive homogeneity of the local, AVE and maximin criteria (see e.g. Pronzato and Pázman, 2013, Sects. 8.1 and 8.2) and, typically, the optimal design for nonlinear model maximizes one of them.

1.4 The cutting plane method

The method of cutting planes (Kelley, 1960) can be applied when optimizing a concave criterion, see also Sect. 9.5 in Pronzato and Pázman (2013).

Let $\varphi : \mathbf{z} \to \varphi(\mathbf{z})$ be a continuous concave function of a vector variable $\mathbf{z}$ defined on a compact convex set. The optimization problem

$$\mathbf{z}^\star = \arg\max_{\mathbf{z}} \varphi(\mathbf{z}) \tag{1.11}$$

can be solved by the iterative cutting plane algorithm, which follows from the subgradient inequality, see (1.13). Suppose that $\mathbf{z}^{(0)}$ is a starting point and $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(i)}$ are results of previous iterations. The solution of $(i+1)$-st iteration is computed as follows

$$\mathbf{z}^{(i+1)} = \arg\max_{\mathbf{z}} \min_{j=0,1,\ldots,i} \left[\varphi\left(\mathbf{z}^{(j)}\right) + \nabla^\top \varphi\left(\mathbf{z}^{(j)}\right)\left(\mathbf{z} - \mathbf{z}^{(j)}\right)\right], \tag{1.12}$$

where $\nabla\varphi\left(\mathbf{z}^{(j)}\right)$ is an arbitrary subgradient of $\varphi(\cdot)$ at $\mathbf{z}^{(j)}$ (one may here also use the term supergradient, since φ is concave function). The problem (1.12) is a linear programming (LP) problem and the standard LP algorithms may be used. This is the main idea of the cutting plane method presented in Sect. 9.5 of Pronzato and Pázman (2013). Algorithm stops when for given $\epsilon > 0$ the inequality

$$\min_{j=0,1,\ldots,i}\left[\varphi\left(\mathbf{z}^{(j)}\right)+\nabla^{\top}\varphi\left(\mathbf{z}^{(j)}\right)\left(\mathbf{z}^{(i+1)}-\mathbf{z}^{(j)}\right)\right]-\max_{l=0,\ldots,i+1}\varphi\left(\mathbf{z}^{(l)}\right)<\epsilon$$

is satisfied, and then $\arg\max_{\mathbf{z}\in\left\{\mathbf{z}^{(0)},\mathbf{z}^{(1)},\ldots,\mathbf{z}^{(i+1)}\right\}}\varphi(\mathbf{z})$ is the computed solution of the problem (1.11).

1.4.1 A subgradient of a concave function

Here we summarize some important properties of subgradients of a concave function (see Pronzato and Pázman (2013) Sect. 9.5 and Appendix A), which will be useful in the next parts of the thesis. For more detailed explanation we refer to Sects. 3.1.5–3.1.6 in Nesterov (2004) or Sect. 23 in Rockafellar (1970) (where subgradients of a convex function are considered).

The subgradient of a concave function φ at point $\tilde{\mathbf{z}}$, the element of the domain of φ, is any vector $\nabla\varphi(\tilde{\mathbf{z}})$ satisfying the following inequality for any $\mathbf{z}$ in the domain of φ:

$$\varphi(\mathbf{z})\leq\varphi(\tilde{\mathbf{z}})+\nabla^{\top}\varphi(\tilde{\mathbf{z}})(\mathbf{z}-\tilde{\mathbf{z}}). \tag{1.13}$$

The inequality (1.13) is known as subgradient inequality. When the function φ is differentiable at $\tilde{\mathbf{z}}$, then there is only one subgradient, which is, moreover, equal to the gradient of φ at $\tilde{\mathbf{z}}$, see Lemma 3.1.7 in Nesterov (2004).

According to Theorem 23.4 in Rockafellar (1970), the subgradient of φ does exist at $\tilde{\mathbf{z}}$ if and only if $\tilde{\mathbf{z}}$ is an element of interior of the domain of φ. When applying the cutting plane method in the optimal experimental design (supposing that $\mathcal{X}$ is finite), there is a need to express a subgradient at some design $\tilde{\xi}$. Although we accept only such experimental designs ξ which are non-negative in each component and summing to one (i.e. $\xi\in\Xi$, a subset of $\mathbb{R}^{\mathrm{card}(\mathcal{X})}$ with an empty interior), generally, the natural domain of the criterion function $\mathbf{w}\mapsto\phi(\mathbf{w},\boldsymbol{\theta})$ (the maximal set of vectors $\mathbf{w}$ for which the function ϕ is defined) is much more larger and hence the necessary and sufficient condition for existence of a subgradient is not very restrictive.

In the next lemma we summarize some well-known properties of subgradients. For more general results see Sect. 3.1.6 in Nesterov (2004).

Lemma 1.4. *Let $\varphi(\cdot),\varphi_1(\cdot)$ and $\varphi_2(\cdot)$ be concave functions of a vector variable $\mathbf{z}$ with the same domain. Suppose that theirs subgradients at $\tilde{\mathbf{z}}$ do exist and are denoted by $\nabla\varphi(\tilde{\mathbf{z}}),\nabla\varphi_1(\tilde{\mathbf{z}}),\nabla\varphi_2(\tilde{\mathbf{z}})$. Then*

a) $[a\nabla\varphi(\widetilde{\mathbf{z}})]$ *is a subgradient of the function* $[a\varphi(\mathbf{z})]$ *at* $\widetilde{\mathbf{z}}$ *for any* $a \geq 0$,

b) $[a_1\nabla\varphi_1(\widetilde{\mathbf{z}}) + a_2\nabla\varphi_2(\widetilde{\mathbf{z}})]$ *is a subgradient of* $\varphi(\mathbf{z}) = a_1\varphi_1(\mathbf{z}) + a_2\varphi_2(\mathbf{z})$ *at* $\widetilde{\mathbf{z}}$ *for any* $a_1, a_2 \geq 0$,

c) define the set $\mathcal{I}(\widetilde{\mathbf{z}}) = \{i \in \{1,2\}, \varphi_i(\widetilde{\mathbf{z}}) = \min\{\varphi_1(\widetilde{\mathbf{z}}), \varphi_2(\widetilde{\mathbf{z}})\}\}$, *then* $\nabla\varphi_{i^\star}(\widetilde{\mathbf{z}})$ *is a subgradient of* $\varphi(\mathbf{z}) = \min\{\varphi_1(\mathbf{z}), \varphi_2(\mathbf{z})\}$ *at* $\widetilde{\mathbf{z}}$, *where* $i^\star$ *is any index from* $\mathcal{I}(\widetilde{\mathbf{z}})$,

d) let $\varrho : \mathbf{z} \times \boldsymbol{\theta} \mapsto \varrho(\mathbf{z}, \boldsymbol{\theta})$ *be concave function in* $\mathbf{z}$ *for any* $\boldsymbol{\theta} \in \Theta$ *(*$\boldsymbol{\theta}$ *here represents a random variable or a random vector), then* $E[\nabla\varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta})]$ *is a subgradient of* $\varphi(\mathbf{z}) = E[\varrho(\mathbf{z}, \boldsymbol{\theta})]$ *at* $\widetilde{\mathbf{z}}$.

Proof.

a) Multiplying the inequality (1.13) by a one obtains the subgradient inequality for $[a\varphi(\cdot)]$ at $\widetilde{\mathbf{z}}$.

b) Using the statement a) of this lemma and summing up the subgradient inequalities for $a_1\varphi_1(\cdot)$ and $a_2\varphi_2(\cdot)$ one obtains the subgradient inequality for the sum $[a_1\varphi_1(\cdot) + a_2\varphi_2(\cdot)]$.

c) The function $\varphi(\mathbf{z}) = \min\{\varphi_1(\mathbf{z}), \varphi_2(\mathbf{z})\}$ is concave in $\mathbf{z}$ and according to (1.13) for any $i^\star \in \mathcal{I}(\widetilde{\mathbf{z}})$ one has $\varphi(\mathbf{z}) = \min\{\varphi_1(\mathbf{z}), \varphi_2(\mathbf{z})\} \leq \varphi_{i^\star}(\mathbf{z}) \leq \varphi_{i^\star}(\widetilde{\mathbf{z}}) + \nabla^\top\varphi_{i^\star}(\widetilde{\mathbf{z}})(\mathbf{z} - \widetilde{\mathbf{z}}) = \varphi(\widetilde{\mathbf{z}}) + \nabla^\top\varphi_{i^\star}(\widetilde{\mathbf{z}})(\mathbf{z} - \widetilde{\mathbf{z}})$.

d) In the proof we follow Vandenberghe (2016). According to (1.13) we have $\varrho(\mathbf{z}, \boldsymbol{\theta}) \leq \varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta}) + \nabla^\top\varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta})(\mathbf{z} - \widetilde{\mathbf{z}})$ for any $\mathbf{z}, \widetilde{\mathbf{z}}, \boldsymbol{\theta}$, which implies the subgradient inequality for φ:

$$\begin{aligned}\varphi(\mathbf{z}) = E[\varrho(\mathbf{z}, \boldsymbol{\theta})] &\leq E\left[\varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta}) + \nabla^\top\varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta})(\mathbf{z} - \widetilde{\mathbf{z}})\right] \\ &= E[\varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta})] + E\left[\nabla^\top\varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta})\right](\mathbf{z} - \widetilde{\mathbf{z}}) \\ &= \varphi(\widetilde{\mathbf{z}}) + E\left[\nabla^\top\varrho(\widetilde{\mathbf{z}}, \boldsymbol{\theta})\right](\mathbf{z} - \widetilde{\mathbf{z}}).\end{aligned}$$

□

Chapter 2

Experimental design via linear programming

This chapter is based on the paper Burclová and Pázman (2016a), where we elaborated an idea of computing optimal experimental designs via linear programming (LP). In the paper we mostly considered the classical linear regression model (1.4), but the reformulation of results of the paper for locally optimal designs in the classical nonlinear regression model (1.5) (or even in a generalized regression model, see Sect. 4.2.1) is straightforward. Since the thesis is focused on nonlinear models, here we present those results of Burclová and Pázman (2016a) which are applicable to local (1.8) and AVE (1.10) optimality criteria.

In Sect. 2.1 we mention the expressions presented in Sect. 9.5.3 of Pronzato and Pázman (2013) for computation of E-, $\mathbf{c}$-, and G-optimal designs. By the use of several results of matrix algebra (c.f. e.g. Harville, 2008) we reformulate the criteria of D-, A-, and E_k-optimality (see Def. 1.2) in Sect. 2.2 for computation of locally and in Sect. 2.4 of AVE optimal designs in nonlinear models. In Sect. 2.3 we present the algorithm based on the cutting plane method and we formulate two more complicated experimental design problems which we could solve directly. Section 2.5 provides some bibliographic remarks and Sect. 2.6 concludes.

Throughout this chapter we will always suppose that the design space $\mathcal{X}$ is finite and hence the design ξ can be considered as a $\operatorname{card}(\mathcal{X})$-dimensional vector with nonnegative components summing to one. We will use the notation $\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right) = \frac{\partial\eta(\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$, where $\eta(\mathbf{x},\boldsymbol{\theta})$ is the regression function in the classical nonlinear regression model (1.5). It follows that the elementary information matrix (1.6) can be expressed as $M\left(\mathbf{x},\boldsymbol{\theta}^0\right) = \mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)\mathbf{f}^\top\left(\mathbf{x},\boldsymbol{\theta}^0\right)$.

2.1 Reformulation of criteria of E-, c-, and G-optimality

In Sect. 9.5.3 of Pronzato and Pázman (2013) one can find that the local criteria of E- and $\mathbf{c}$-optimality can be rewritten into a form

$$\phi\left(\xi,\boldsymbol{\theta}^0\right)=\min_{\mathbf{u}\in\mathcal{U}}\sum_{\mathbf{x}\in\mathcal{X}}\mathbf{u}^\top M\left(\mathbf{x},\boldsymbol{\theta}^0\right)\mathbf{u}\xi\left(\mathbf{x}\right),$$

where in the case of E-optimality $\mathcal{U}=\left\{\mathbf{u}\in\mathbb{R}^m:\|\mathbf{u}\|_{\ell(2)}=1\right\}$ and in the case of $\mathbf{c}$-optimality $\mathcal{U}=\left\{\mathbf{u}\in\mathbb{R}^m:\mathbf{u}^\top\mathbf{c}=1\right\}$. For the criterion of G-optimality one obtains $\phi_G\left(\xi,\boldsymbol{\theta}^0\right)=\min_{\mathbf{x}\in\mathcal{X}}\min_{\mathbf{u}:\mathbf{u}^\top\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)=1}\mathbf{u}^\top M\left(\mathbf{x},\boldsymbol{\theta}^0\right)\mathbf{u}\xi\left(\mathbf{x}\right)$.

This means that the problem of finding locally optimal design $\xi^\star$ for these criteria is an "infinite-dimensional" LP problem:

$$\begin{aligned}
&\max_{\xi,t}\left(\mathbf{0}^\top,1\right)\begin{pmatrix}\xi\\ t\end{pmatrix}\\
\text{s.t.}\ \sum_{\mathbf{x}\in\mathcal{X}}\mathbf{u}^\top M\left(\mathbf{x},\boldsymbol{\theta}^0\right)\mathbf{u}\xi\left(\mathbf{x}\right)&\geq t\ \text{for any}\ \mathbf{u}\in\mathcal{U},\\
\xi\left(\mathbf{x}\right)&\geq 0\ \text{for any}\ \mathbf{x}\in\mathcal{X},\ \text{and}\ \sum_{\mathbf{x}\in\mathcal{X}}\xi\left(\mathbf{x}\right)=1,
\end{aligned}$$

where t is an auxiliary variable and $\xi=(\xi(\mathbf{x});\mathbf{x}\in\mathcal{X})$ is a $\mathrm{card}(\mathcal{X})$-dimensional vector. The algorithm and the modification for G-optimality can be found in Sect. 9.5.3 of Pronzato and Pázman (2013).

2.2 Reformulation of criteria of D-, A-, and E_k-optimality

Let us denote by Ξ^+ the set of such designs ζ for which the information matrix $M\left(\zeta,\boldsymbol{\theta}^0\right)$ is nonsingular, i.e. $\Xi^+=\left\{\zeta\in\Xi:M\left(\zeta,\boldsymbol{\theta}^0\right)\ \text{is nonsingular}\right\}$. Evidently, D-, A-, and E-locally optimal designs belong to Ξ^+, which is not necessarily true for E_k-optimality when $k\geq 2$. As we already mentioned, the following sections are based on the publication Burclová and Pázman (2016a).

Theorem 2.1. *We can write* $\phi_D\left(\xi,\boldsymbol{\theta}^0\right)=\min_{\zeta\in\Xi^+}\sum_{\mathbf{x}\in\mathcal{X}}H_D\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)\xi\left(\mathbf{x}\right)$ *for any* $\xi\in\Xi^+$, *where*

$$H_D\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)=\frac{\det^{1/m}\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]}{m}\mathbf{f}^\top\left(\mathbf{x},\boldsymbol{\theta}^0\right)M^{-1}\left(\zeta,\boldsymbol{\theta}^0\right)\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right).\qquad(2.1)$$

Proof. Take any nonsingular square matrix $S\in\mathbb{R}^{m\times m}$ and denote by $\beta_1,\ldots,\beta_m$ the eigenvalues of $S^\top M\left(\xi,\boldsymbol{\theta}^0\right)S$. Since the geometric mean of positive numbers $\beta_1,\ldots,\beta_m$

is always less than or equal to the arithmetic mean of $\beta_1, \ldots, \beta_m$ (cf. e.g. Steele, 2004, Chap. 2), we obtain

$$\left\{\det\left[S^\top M\left(\xi, \boldsymbol{\theta}^0\right) S\right]\right\}^{1/m} = \left(\prod_{i=1}^m \beta_i\right)^{1/m} \leq \frac{1}{m}\sum_{i=1}^m \beta_i = \frac{1}{m}\operatorname{tr}\left[S^\top M\left(\xi, \boldsymbol{\theta}^0\right) S\right].$$

So

$$\phi_D\left(\xi, \boldsymbol{\theta}^0\right) = \left\{\det\left[M\left(\xi, \boldsymbol{\theta}^0\right)\right]\right\}^{\frac{1}{m}} \leq \frac{\left[\det\left(SS^\top\right)\right]^{\frac{-1}{m}}}{m} \sum_{\mathbf{x}\in\mathcal{X}} \mathbf{f}^\top\left(\mathbf{x}, \boldsymbol{\theta}^0\right) SS^\top \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right) \xi\left(\mathbf{x}\right),$$

and we have only to put $S = M^{-1/2}\left(\zeta, \boldsymbol{\theta}^0\right)$ to obtain on the right hand side the expression in (2.1). If $S = M^{-1/2}\left(\xi, \boldsymbol{\theta}^0\right)$, then $\beta_1 = \ldots = \beta_m = 1$, and the geometric mean is equal to the arithmetic mean, hence we have the equality for $\zeta = \xi$ which proves the theorem. □

Theorem 2.2. *We can write* $\phi_A\left(\xi, \boldsymbol{\theta}^0\right) = \min_{\zeta\in\Xi^+} \sum_{\mathbf{x}\in\mathcal{X}} H_A\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi\left(\mathbf{x}\right)$ *for any* $\xi \in \Xi^+$, *where*

$$H_A\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) = \frac{\left\|M^{-1}\left(\zeta, \boldsymbol{\theta}^0\right) \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\right\|^2_{\ell(2)}}{\left\{\operatorname{tr}\left[M^{-1}\left(\zeta, \boldsymbol{\theta}^0\right)\right]\right\}^2} \tag{2.2}$$

and $\|\cdot\|_{\ell(2)}$ *denotes the Euclidean norm.*

Proof. For any nonsingular matrix $S \in \mathbb{R}^{m\times m}$ and for any $\xi \in \Xi^+$ we obtain from the Cauchy-Schwarz inequality (see Eq. (1.1) after substitution $A \leftarrow S^\top$, $B \leftarrow I$, $M \leftarrow M\left(\xi, \boldsymbol{\theta}^0\right)$) that

$$\left[\operatorname{tr}\left(S\right)\right]^2 \leq \operatorname{tr}\left[M^{-1}\left(\xi, \boldsymbol{\theta}^0\right)\right] \operatorname{tr}\left[SM\left(\xi, \boldsymbol{\theta}^0\right) S^\top\right].$$

So

$$\phi_A\left(\xi, \boldsymbol{\theta}^0\right) = \left\{\operatorname{tr}\left[M^{-1}\left(\xi, \boldsymbol{\theta}^0\right)\right]\right\}^{-1} \leq \frac{\operatorname{tr}\left[SM\left(\xi, \boldsymbol{\theta}^0\right) S^\top\right]}{\left[\operatorname{tr}\left(S\right)\right]^2} = \sum_{\mathbf{x}\in\mathcal{X}} \frac{\left\|S\mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\right\|^2_{\ell(2)}}{\left[\operatorname{tr}\left(S\right)\right]^2} \xi\left(\mathbf{x}\right),$$

and we have only to put $S = M^{-1}\left(\zeta, \boldsymbol{\theta}^0\right)$ to obtain on the right-hand side the expression in (2.2). When $S = M^{-1}\left(\xi, \boldsymbol{\theta}^0\right)$, we obtain equality in the Cauchy-Schwarz inequality and the theorem is proved. □

Theorem 2.3. *Denote by* $\mathbf{u}_1\left[M\left(\xi, \boldsymbol{\theta}^0\right)\right], \ldots, \mathbf{u}_m\left[M\left(\xi, \boldsymbol{\theta}^0\right)\right]$ *the orthonormal eigenvectors of* $M\left(\xi, \boldsymbol{\theta}^0\right)$ *corresponding to* $\lambda_1\left[M\left(\xi, \boldsymbol{\theta}^0\right)\right], \ldots, \lambda_m\left[M\left(\xi, \boldsymbol{\theta}^0\right)\right]$. *We can write* $\phi_{E_k}\left(\xi, \boldsymbol{\theta}^0\right) = \min_{\zeta\in\Xi} \sum_{\mathbf{x}\in\mathcal{X}} H_{E_k}\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi\left(\mathbf{x}\right)$ *for any* $\xi \in \Xi$, *where*

$$H_{E_k}\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) = \left\|P^{(k)}\left(\zeta, \boldsymbol{\theta}^0\right) \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\right\|^2_{\ell(2)}. \tag{2.3}$$

Here $P^{(k)}\left(\zeta,\boldsymbol{\theta}^0\right)$ *is the* k*-dimensional orthogonal projector*

$$P^{(k)}\left(\zeta,\boldsymbol{\theta}^0\right)=\sum_{i=1}^{k}\mathbf{u}_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]\mathbf{u}_i^{\top}\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right],$$

and $\|\cdot\|_{\ell(2)}$ *denotes the Euclidean norm. When the eigenvalue* $\lambda_k\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]$ *has multiplicity* $p>1$*, the choice of the order of the* p *orthonormal eigenvectors corresponding to* $\lambda_k\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]$ *is arbitrary.*

Proof. For a fixed ζ denote $P=P^{(k)}\left(\zeta,\boldsymbol{\theta}^0\right)$. By the definition of $P^{(k)}\left(\zeta,\boldsymbol{\theta}^0\right)$ we have that P is idempotent symmetric matrix. Moreover, for given ξ define the orthogonal matrix $U=\left(\mathbf{u}_1\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right],\ldots,\mathbf{u}_m\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right]\right)$ and the diagonal matrix of eigenvalues $\Lambda=\operatorname{diag}\left\{\lambda_1\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right],\ldots,\lambda_m\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right]\right\}$. Hence one has $M\left(\xi,\boldsymbol{\theta}^0\right)=U\Lambda U^{\top}$. It follows that

$$\begin{aligned}\sum_{\mathbf{x}\in\mathcal{X}}\left\|P\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)\right\|_{\ell(2)}^2\xi\left(\mathbf{x}\right)=\operatorname{tr}\left[PM\left(\xi,\boldsymbol{\theta}^0\right)P\right] &= \operatorname{tr}\left[PU\Lambda U^{\top}P\right]=\operatorname{tr}\left[\Lambda\left(PU\right)^{\top}\left(PU\right)\right]\\ &= \sum_{i=1}^{m}\lambda_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]\left\{\left(PU\right)^{\top}\left(PU\right)\right\}_{ii}\\ &= \sum_{i=1}^{m}\lambda_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]w_i,\end{aligned}$$

where we denoted $w_i=\left\{\left(PU\right)^{\top}\left(PU\right)\right\}_{ii}=\left\|P\mathbf{u}_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]\right\|_{\ell(2)}^2$. Since U is orthogonal matrix, i.e. $UU^{\top}=U^{\top}U=I$, we have

$$k=\operatorname{tr}\left(P\right)=\operatorname{tr}\left(P^{\top}P\right)=\operatorname{tr}\left(P^{\top}PUU^{\top}\right)=\sum_{i=1}^{m}\left\{\left(PU\right)^{\top}\left(PU\right)\right\}_{ii}=\sum_{i=1}^{m}w_i.$$

Further, $w_i\in[0,1]$ since $0\leq\left\|P\mathbf{u}_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]\right\|_{\ell(2)}^2\leq\left\|\mathbf{u}_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]\right\|_{\ell(2)}^2=1$. So, using the inequalities $\lambda_1\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]\leq\ldots\leq\lambda_m\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]$, we can see that the expression $\sum_{i=1}^{m}\lambda_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]w_i$ is minimized exactly when $w_1=\ldots=w_k=1$ and $w_{k+1}=\ldots=w_m=0$, hence

$$\sum_{\mathbf{x}\in\mathcal{X}}\left\|P\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)\right\|_{\ell(2)}^2\xi\left(\mathbf{x}\right)=\sum_{i=1}^{m}\lambda_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]w_i\geq\sum_{i=1}^{k}\lambda_i\left[M\left(\zeta,\boldsymbol{\theta}^0\right)\right]=\phi_{E_k}\left(\xi,\boldsymbol{\theta}^0\right),\tag{2.4}$$

where on the left-hand side is the expression from (2.3). In the particular case when $P=P^{(k)}\left(\xi,\boldsymbol{\theta}^0\right)=\sum_{j=1}^{k}\mathbf{u}_j\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right]\mathbf{u}_j^{\top}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right]$, we have

$$w_i=\left\|P^{(k)}\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u}_i\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right]\right\|_{\ell(2)}^2=\begin{cases}\left\|\mathbf{u}_i\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right]\right\|_{\ell(2)}^2 & \text{if } i\leq k,\\ 0 & \text{if } i>k,\end{cases}=\begin{cases}1 & \text{if } i\leq k,\\ 0 & \text{if } i>k,\end{cases}$$

and hence $\sum_{\mathbf{x}\in\mathcal{X}} \left\| P^{(k)}\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)\right\|^2_{\ell(2)}\xi\left(\mathbf{x}\right) = \sum_{i=1}^{k}\lambda_i\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right] = \phi_{E_k}\left(\xi,\boldsymbol{\theta}^0\right)$, which together with (2.4) yields the statement of the theorem. From the presented proof it is evident that the choice of the order of eigenvectors corresponding to an eigenvalue with multiplicity greater than 1 is irrelevant, as long as these eigenvectors are orthonormal. □

From the reformulated criteria functions in Theorems 2.1–2.3 it is easy to see that the criteria of D-, A-, and E_k optimality from Def. 1.2 are concave functions of the design ξ and that they are positively homogeneous.

Let us use the common notation $H\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)$ for $H_D\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)$, $H_A\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)$, or $H_{E_k}\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)$ from Eqs. (2.1)–(2.3). The reformulations in Theorems 2.1–2.3 allow us to find a locally optimal experimental design $\xi^\star = \arg\max_{\xi\in\Xi}\phi\left(\xi,\boldsymbol{\theta}^0\right)$ via an "infinite-dimensional" LP problem:

$$\begin{aligned} &\max\left(\mathbf{0}^\top,1\right)\begin{pmatrix}\xi\\ t\end{pmatrix}\\ \text{s.t.}\ &\sum_{\mathbf{x}\in\mathcal{X}} H\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)\xi\left(\mathbf{x}\right)\geq t \text{ for any } \zeta\in\Xi^*, \\ &\qquad\qquad \xi\left(\mathbf{x}\right)\geq 0 \text{ for any } \mathbf{x}\in\mathcal{X}, \text{ and } \sum_{\mathbf{x}\in\mathcal{X}}\xi\left(\mathbf{x}\right)=1, \end{aligned} \tag{2.5}$$

where t is an auxiliary variable and $\xi = (\xi(\mathbf{x}); \mathbf{x}\in\mathcal{X})$ is a $\mathrm{card}(\mathcal{X})$-dimensional vector. Here Ξ^* denotes either Ξ or Ξ^+, depending on the criterion considered.

Remark 2.4. It follows directly from the proofs of Theorems 2.1–2.3 that one can write for a fixed $\xi\in\Xi^+$

$$\phi_A\left(\xi,\boldsymbol{\theta}^0\right) = \min_{B\in\mathcal{B}}\sum_{\mathbf{x}\in\mathcal{X}}\left\{\frac{\left\|B\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)\right\|^2_{\ell(2)}}{\left[\mathrm{tr}\left(B\right)\right]^2}\right\}\xi\left(\mathbf{x}\right),$$

where $\mathcal{B}$ is any set of nonsingular matrices in $\mathbb{R}^{m\times m}$ containing $M^{-1}\left(\xi,\boldsymbol{\beta}^0\right)$. To ensure the validity of the statement for any $\xi\in\Xi^+$, the set $\mathcal{B} = \left\{M^{-1}\left(\zeta,\boldsymbol{\beta}^0\right) : \zeta\in\Xi^+\right\}$, used also in Theorem 2.2, is the smallest of such sets. In the same way one obtains for D-optimality, see Theorem 2.1:

$$\phi_D\left(\xi,\boldsymbol{\theta}^0\right) = \min_{B\in\mathcal{B}^+}\sum_{\mathbf{x}\in\mathcal{X}}\frac{1}{m}\left\{\frac{\mathbf{f}^\top\left(\mathbf{x},\boldsymbol{\theta}^0\right)B\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)}{\det^{1/m}\left(B\right)}\right\}\xi\left(\mathbf{x}\right),$$

where $\mathcal{B}^+$ is any set of positive definite symmetric matrices in $\mathbb{R}^{m\times m}$ including $M^{-1}\left(\xi,\boldsymbol{\theta}^0\right)$. Similarly, in Theorem 2.3 we could minimize over any set $\mathcal{P}$ of k-dimensional orthogonal projectors containing $P^{(k)}\left(\xi,\boldsymbol{\theta}^0\right)$:

$$\phi_{E_k}\left(\xi,\boldsymbol{\theta}^0\right) = \min_{P\in\mathcal{P}}\sum_{\mathbf{x}\in\mathcal{X}}\left\|P\mathbf{f}\left(\mathbf{x},\boldsymbol{\theta}^0\right)\right\|^2_{\ell(2)}\xi\left(\mathbf{x}\right).$$

This follows from the proof of Theorem 2.3, where we can substitute any k-dimensional orthogonal projector P for $P^{(k)}\left(\zeta, \boldsymbol{\theta}^0\right)$.

2.3 Algorithm and example

In this section we use the reformulated optimality criteria from Theorems 2.1–2.3 to compute locally optimal designs. For the maximization of $\phi\left(\cdot, \boldsymbol{\theta}^0\right)$, i.e. for solving the LP problem (2.5), we apply a modification of the relaxation method (Shimizu and Aiyoshi, 1980) as presented in Sect. 9.5.3 of Pronzato and Pázman (2013) and in Pázman and Pronzato (2014), where its relation to the cutting plane method (Kelley, 1960) is also shown. The presented algorithm is iterative and we use an LP solver at each iteration.

Algorithm 2.5.

0. i Take the starting design $\xi^{(0)} \in \mathbb{R}^{\mathrm{card}(\mathcal{X})}$, s.t. $\sum_{\mathbf{x}\in\mathcal{X}} \xi^{(0)}(\mathbf{x}) = 1$ and $\xi^{(0)}(\mathbf{x}) \geq 0 \quad \forall\, \mathbf{x} \in \mathcal{X}$.

   ii Choose the accuracy $\epsilon > 0$.

   iii Set $\Xi^{(0)} = \emptyset$.

   iv Set $\xi' = \xi^{(0)}$.

   v Set $i = 1$.

1. Let $\Xi^{(i)} = \Xi^{(i-1)} \cup \left\{\xi^{(i-1)}\right\}$.

2. LP problem: find $\left(\xi^{(i)}, t^{(i)}\right)$ which maximizes $t^{(i)}$ under the constraints

$$
\begin{aligned}
& t^{(i)} > 0,\ \xi^{(i)}(\mathbf{x}) \geq 0 \quad \forall\, \mathbf{x} \in \mathcal{X}, \\
& \sum_{\mathbf{x}\in\mathcal{X}} \xi^{(i)}(\mathbf{x}) = 1, \\
& \sum_{\mathbf{x}\in\mathcal{X}} H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi^{(i)}(\mathbf{x}) \geq t^{(i)} \quad \forall \zeta \in \Xi^{(i)}.
\end{aligned}
$$

3. i Set $\xi' = \arg\max_{\xi\in\left\{\xi', \xi^{(i)}\right\}} \phi\left(\xi, \boldsymbol{\theta}^0\right)$.

   ii If $t^{(i)} - \phi\left(\xi', \boldsymbol{\theta}^0\right) < \epsilon$, return $\xi^\star_\epsilon = \xi'$ as an ϵ-optimal design and stop.

   iii Else set $i \leftarrow i + 1$ and continue from Step 1.

A simple geometric interpretation of the algorithm follows from the fact that $\min_{\zeta\in\Xi^{(i)}} \sum_{\mathbf{x}\in\mathcal{X}} H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x})$ is an upper piecewise linear approximation of $\phi\left(\xi, \boldsymbol{\theta}^0\right)$. Increasing i, the set $\Xi^{(i)} \subseteq \Xi$ becomes larger, and the approximation improves.

On the other hand, when the number of iteration i is small, the information matrix $M\left(\xi^{(i)}, \boldsymbol{\theta}^0\right)$ associated with the solution of the LP problem in the i-th iteration can be ill-conditioned or even singular. In such case, there could be difficulties with inverse matrices when evaluating the function $H_D\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right)$ or $H_A\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right)$ in the $(i+1)$-st iteration for $\zeta = \xi^{(i)}$. However, as justified in Remark 2.4, it is possible to use any symmetric positive definite matrix as a substitute for $M\left(\zeta, \boldsymbol{\theta}^0\right)$. Therefore, if the matrix $M\left(\zeta, \boldsymbol{\theta}^0\right)$ is ill-conditioned or singular, we replace it by the matrix $M\left(\zeta, \boldsymbol{\theta}^0\right) + \beta I$, where β is a small positive number. Note that it is also possible to take $\Xi^{(0)}$ as a nonempty set containing $s \geq 1$ initial designs. If s or i is large, then the solution $\xi^{(i)}$ of the LP problem in the i-th iteration will be more probably associated with a nonsingular and well-conditioned information matrix, because we expect a better approximation of the criterion $\phi\left(\cdot, \boldsymbol{\theta}^0\right)$ in Step 2 of the algorithm. Note that the problem of a singular information matrix does not appear in the case of E_k-optimality.

The stopping rule used in the above algorithm follows from the upper and lower bounds for $\max_{\xi \in \Xi} \phi\left(\xi, \boldsymbol{\theta}^0\right)$:

$$\phi\left(\xi', \boldsymbol{\theta}^0\right) \leq \max_{\xi \in \Xi} \phi\left(\xi, \boldsymbol{\theta}^0\right) \leq t^{(i)}. \tag{2.6}$$

The first inequality is obvious. Notice that the solution of the LP problem in the i-th iteration is $t^{(i)} = \max_{\xi \in \Xi} \min_{\zeta \in \Xi^{(i)}} \sum_{\mathbf{x} \in \mathcal{X}} H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x})$, while $\max_{\xi \in \Xi} \phi\left(\xi, \boldsymbol{\theta}^0\right) = \max_{\xi \in \Xi} \min_{\zeta \in \Xi} \sum_{\mathbf{x} \in \mathcal{X}} H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x})$, and $\Xi \supseteq \Xi^{(i)}$. This yields the second inequality.

More standard stopping rules are based on the equivalence theorem (Kiefer and Wolfowitz, 1959; Kiefer, 1974). Let ϵ_{stop} be a small nonnegative number. An iterative algorithm will stop if $d(\xi') < \epsilon_{\text{stop}}$, where we have

$$d\left(\xi'\right) = \left| \max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}^\top\left(\mathbf{x}, \boldsymbol{\theta}^0\right) M^{-1}\left(\xi', \boldsymbol{\theta}^0\right) \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right) - m \right|$$

for D-optimality and

$$d\left(\xi'\right) = \left| \max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}^\top\left(\mathbf{x}, \boldsymbol{\theta}^0\right) M^{-2}\left(\xi', \boldsymbol{\theta}^0\right) \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right) - \operatorname{tr}\left[M^{-1}\left(\xi', \boldsymbol{\theta}^0\right)\right] \right|,$$

for A-optimality, see e.g. Kiefer (1974, 1975). According to Harman (2004), a similar stopping rule for E_k-optimality is

$$d\left(\xi'\right) = \left| \max_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^{k} \left\{ \mathbf{f}^\top\left(\mathbf{x}, \boldsymbol{\theta}^0\right) \mathbf{u}_i \left[M\left(\xi', \boldsymbol{\theta}^0\right)\right] \right\}^2 - \phi_{E_k}\left(\xi', \boldsymbol{\theta}^0\right) \right|,$$

which can be used only if $\lambda_k(\xi') < \lambda_{k+1}(\xi')$. When $\lambda_k(\xi')$ has multiplicity greater than one, the equivalence theorem could be still based on the directional derivative of ϕ_{E_k}, which exists due to the concavity of $\phi_{E_k}\left(\cdot, \boldsymbol{\theta}^0\right)$ but is difficult to compute.

As mentioned in Sect. 9.5.3 in Pronzato and Pázman (2013), the cutting plane method can lead to instabilities when the number of elements in the design space $\mathcal{X}$ is

large (see Nesterov, 2004; Bonnans et al., 2006). Then the level method (see Nesterov, 2004, Chap. 3.3.2 or Pronzato and Pázman, 2013, Sect. 9.5.3), which employs both linear and quadratic programming, can be used.

For the time being we are not able to prove the convergence of the presented algorithm, so we have to consider our method as a search method with a reliable stopping rule (2.6). But it is useful to present the heuristic which is behind this search.

The algorithm solves the maximin problem $\xi^\star = \arg\max_{\xi\in\Xi} \min_{\zeta\in\Xi} Q(\zeta,\xi)$, where $Q(\zeta,\xi) = \sum_{\mathbf{x}\in\mathcal{X}} H\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)\xi(\mathbf{x})$ is a known "pay-off" function of the vector variables ζ and ξ. We remind here that $\mathcal{X}$ is a finite set, so ζ and ξ are $\mathrm{card}(\mathcal{X})$-dimensional vectors. Shimizu and Aiyoshi (1980) studied a similar maximin problem, but with a pay-off function satisfying specific properties. They proved the convergence of a relaxation procedure to their maximin problem, which leads to Algorithm 2.5 presented above. Namely, their approach in terms of our problem involves in the i-th iteration the computation of the "relaxed problem"

$$\max_{t^{(i)}\in\mathbb{R}^+,\, \xi^{(i)}\in\Xi} t^{(i)} \quad \text{under the inequalities} \quad Q\left(\zeta,\xi^{(i)}\right) \geq t^{(i)} \quad \forall\zeta\in\Xi^{(i)}.$$

For our "pay-off" function this is an LP problem (see Step 2 of Algorithm 2.5). Shimizu and Aiyoshi (1980) constructed their set $\Xi^{(i+1)}$ by adding the "worst" ζ in each step, i.e. they chose

$$\zeta_{\mathrm{worst}} \in \arg\min_{\zeta\in\Xi^*} Q\left(\zeta,\xi^{(i)}\right). \tag{2.7}$$

We do in fact the same, since, as follows from the proofs of Theorems 2.1–2.3, in our case $\zeta_{\mathrm{worst}} = \xi^{(i)}$. So we do not need to solve the possibly complicated minimization problem (2.7) and we simply add $\xi^{(i)}$ to $\Xi^{(i)}$ to obtain $\Xi^{(i+1)}$.

Unfortunately, some of the assumptions for the convergence of the procedure formulated in Shimizu and Aiyoshi (1980) are not satisfied here. Although the set Ξ is compact, which is required by Shimizu and Aiyoshi (1980), once the computed $\xi^{(i)}$ belongs to the set $\Xi - \Xi^+$, the expression $H\left(\xi^{(i)},\mathbf{x},\boldsymbol{\theta}^0\right)$ is not defined in the case of D- or A-optimality, and we have to use a regularization in accordance with Remark 2.4, see the discussion after Algorithm 2.5. This violates the assumptions of Shimizu and Aiyoshi (1980). In the case of E_k optimality, the function $H_{E_k}\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)$ is not continuous in the variable ζ, which is required for their proof of convergence. This discontinuity appears when the rank of $M\left(\zeta,\boldsymbol{\theta}^0\right)$ changes with a slight modification of ζ.

In the example below we apply the algorithm to the model which was already studied in Atkinson et al. (1993), so we were able to check the accordance of our results with known optimal designs. The computations were performed in Matlab computing environment and we used the simplex method to solve the LP problems.

Example 2.6. Consider the nonlinear regression model of Atkinson et al. (1993):

$$\eta(x,\boldsymbol{\theta}) = \theta_1\left[\exp(-\theta_2 x) - \exp(-\theta_3 x)\right],\ x \in \mathbb{R}^+,\ \boldsymbol{\theta} = (\theta_1,\theta_2,\theta_3)^\top.$$

We use Algorithm 2.5 to compute locally D- and E_1-optimal designs for the nominal parameter value $\boldsymbol{\theta}^0 = (21.8, 0.05884, 4.298)^\top$ (used in Atkinson et al., 1993). We take the finite design space $\mathcal{X} = \{0.001, 0.002, \ldots, 23.999, 24.000\}$ consisting of 24,000 points and $\epsilon = 10^{-7}$. The starting design $\xi^{(0)}$ allocates the mass 1/3 to points $x \in \{0.2, 1, 23\}$ and zero mass to $x \notin \{0.2, 1, 23\}$ (see Pázman and Pronzato, 2014, Example 3). The computed ϵ-optimal designs are given in Table 2.1 and performances of these designs correspond to Atkinson et al. (1993) (D-optimality) and Pázman and Pronzato (2014) (E_1-optimality). △

ϕ	$\xi_\epsilon^\star$	$\phi\left(\xi_\epsilon^\star, \boldsymbol{\theta}^0\right)$	iter.	time	$d(\xi_\epsilon^\star)$
D	$\begin{Bmatrix} 0.229 & 1.389 & 18.417 \\ 0.3333 & 0.3333 & 0.3333 \end{Bmatrix}$	11.7388	46	66.52 s	$7.42 \cdot 10^{-5}$
E_1	$\begin{Bmatrix} 0.169 & 1.394 & 23.402 \\ 0.1993 & 0.6623 & 0.1384 \end{Bmatrix}$	0.3163	26	30.21 s	$2.63 \cdot 10^{-5}$

Table 2.1: Numerical results in Example 2.6: locally ϵ-optimal designs for D- and E_1-optimality (column 2) and corresponding criterion values (column 3) rounded to four decimal digits; number of iterations (column 4) and computational time (column 5) required until the algorithm stopped; the value of the stopping rule $d(\xi_\epsilon^\star)$ based on the equivalence theorem (column 6).

The long computing times, which depend mainly on the dimension of the design space $\mathcal{X}$, is a potential weakness of the presented LP method. On the other hand, one may also use the algorithm several times, starting the computation with a sparse initial design space $\mathcal{X}$. After the algorithm has evaluated support points of the optimal design, one may add some design points adjacent to these support points and repeat the computation on a modified design space (see Pázman and Pronzato, 2014, Example 3). Nevertheless, the LP approach may not be the method of choice for a "simple" problem such as that in Example 2.6, but it can be useful when requiring optimal designs for complicated scenarios such as those indicated in the remainder of this section.

2.3.1 Computation of criterion robust design

The criteria of E_k optimality of Harman (2004) play a special role in the experimental design. They appear in the definition of Schur ordering of designs: we say that the

design ξ is locally not worse than the design ζ with respect to the Schur ordering of designs if $\phi_{E_k}\left(\xi,\boldsymbol{\theta}^0\right) \geq \phi_{E_k}\left(\zeta,\boldsymbol{\theta}^0\right)$ for all $k=1,\ldots,m$, see e.g. Harman (2008) for some results on Schur optimal designs.

Another application of the class of E_k optimality criteria is related to the computation of "criterion robust" design within the class of orthogonally invariant criteria $\mathcal{O}$ (see Filová and Harman, 2013 for the term criterion robust design). The class $\mathcal{O}$ involves such criteria functions $\phi\left(\xi,\boldsymbol{\theta}^0\right)$ which are concave and positively homogeneous in ξ and, moreover, are orthogonally invariant in the sense that $\phi\left(\xi,\boldsymbol{\theta}^0\right) = \bar{\phi}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right] = \bar{\phi}\left[Q^\top M\left(\xi,\boldsymbol{\theta}^0\right) Q\right]$ for any orthogonal matrix Q. We are interested in computation of criterion robust design $\xi^\star_{\text{ef}}$ which is maximin efficient with respect to the class $\mathcal{O}$, i.e.

$$\xi^\star_{\text{ef}} = \arg\max_{\xi\in\Xi}\min_{\phi\in\mathcal{O}}\left[\frac{\phi\left(\xi,\boldsymbol{\theta}^0\right)}{\max_{\nu\in\Xi}\phi\left(\nu,\boldsymbol{\theta}^0\right)}\right], \tag{2.8}$$

where the ratio $\frac{\phi\left(\xi,\boldsymbol{\theta}^0\right)}{\max_{\nu\in\Xi}\phi\left(\nu,\boldsymbol{\theta}^0\right)}$ denotes the ϕ-efficiency of the design ξ at $\boldsymbol{\theta}^0$.

Harman (2004) proved an important simplification in the computation of criterion robust design (2.8), in particular

$$\xi^\star_{\text{ef}} = \arg\max_{\xi\in\Xi}\min_{1\leq k\leq m}\left[\frac{\phi_{E_k}\left(\xi,\boldsymbol{\theta}^0\right)}{\max_{\nu\in\Xi}\phi_{E_k}\left(\nu,\boldsymbol{\theta}^0\right)}\right],$$

i.e. it is sufficient to consider designs that are maximin efficient in the (finite) class of all E_k-optimality criteria. But even this problem is computationally difficult, mainly because the E_k-optimality criteria are not generally differentiable. We showed in Burclová and Pázman (2016a) that we can solve this problem by the LP technique. First, using Theorem 2.3, we compute $E_k\left(opt,\boldsymbol{\theta}^0\right) = \max_{\nu\in\Xi}\phi_{E_k}\left(\nu,\boldsymbol{\theta}^0\right)$ for all k (see Algorithm 2.5), and then we can formulate another "infinite-dimensional" LP problem:

$$\begin{aligned}
&\max_{\xi,t}\left(\mathbf{0}^\top,1\right)\begin{pmatrix}\xi\\ t\end{pmatrix}\\
\text{s.t.}\quad &\sum_{\mathbf{x}\in\mathcal{X}}\frac{H_{E_k}\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)}{E_k\left(opt,\boldsymbol{\theta}^0\right)}\xi\left(\mathbf{x}\right) \;\geq\; t \text{ for any } \zeta\in\Xi \text{ and for every } k\in\{1,\ldots,m\},\\
&\xi\left(\mathbf{x}\right) \;\geq\; 0 \text{ for any } \mathbf{x}\in\mathcal{X}, \text{ and } \sum_{\mathbf{x}\in\mathcal{X}}\xi\left(\mathbf{x}\right)=1.
\end{aligned}$$

In order to compute the criterion robust design, Algorithm 2.5 needs to be modified in Step 2. Actually, the constraints in the LP problem will be:

$$t^{(i)}>0,\ \xi^{(i)}\left(\mathbf{x}\right)\geq 0 \quad \forall\,\mathbf{x}\in\mathcal{X},$$

$$\sum_{\mathbf{x}\in\mathcal{X}}\xi^{(i)}\left(\mathbf{x}\right)=1,$$

$$\sum_{\mathbf{x}\in\mathcal{X}}\frac{H_{E_k}\left(\zeta,\mathbf{x},\boldsymbol{\theta}^0\right)}{E_k\left(opt,\boldsymbol{\theta}^0\right)}\xi^{(i)}\left(\mathbf{x}\right)\geq t^{(i)} \quad \forall\zeta\in\Xi^{(i)} \text{ and } \quad \forall\, k=1,\ldots,m.$$

In the paper Burclová and Pázman (2016a) we illustrated the algorithm on the quadratic regression model on one-, two-, three- and four-dimensional cube. The E_k-optimal and criterion robust designs on these models were studied in Harman (2004); Filová and Harman (2013) analytically or using methods of semidefinite programming in Filová et al. (2012).

2.3.2 Computation of D-optimal design conditioned by a prescribed level of A-optimality

It is not difficult to see that in the considered LP problems solved in Step 2 of Algorithm 2.5 we can simply include some additional constraints linear in ξ, say, a cost constraint $\sum_{\mathbf{x}\in\mathcal{X}} C(\mathbf{x})\,\xi(\mathbf{x}) = C$, where $C(\mathbf{x})$ is the cost of an observation at $\mathbf{x}$, and C is proportional to the total cost allowed for the whole experiment. What is less evident is that we can combine optimality criteria. For instance, when we want to obtain a design which maximizes the D-optimality criterion and retains the value of A-optimality criterion greater than or equal to some prescribed level a, we have to solve the "infinite-dimensional" LP problem:

$$\begin{aligned} &\max_{\xi,t} \left(\mathbf{0}^\top, 1\right) \begin{pmatrix} \xi \\ t \end{pmatrix} \\ \text{s.t.} \quad &\sum_{\mathbf{x}\in\mathcal{X}} H_D\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x}) \;\geq\; t \text{ for any } \zeta \in \Xi^+, \\ &\sum_{\mathbf{x}\in\mathcal{X}} H_A\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x}) \;\geq\; a \text{ for any } \zeta \in \Xi^+, \\ &\qquad\qquad \xi(\mathbf{x}) \;\geq\; 0 \text{ for any } \mathbf{x} \in \mathcal{X}, \text{ and } \sum_{\mathbf{x}\in\mathcal{X}} \xi(\mathbf{x}) = 1. \end{aligned}$$

This problem can be solved by Algorithm 2.5 with a modification in the constraints of the LP problem and in the stopping rule. The modified algorithm and an illustrative example for linear model is presented in Burclová and Pázman (2016a).

2.4 Reformulation of AVE criteria

The reformulation of expressions in Theorems 2.1–2.3 in terms of AVE optimality criteria (1.10) was also presented in Burclová and Pázman (2016a).

Theorem 2.7. *Denote* $\Xi_\Theta = \{\xi : M(\xi, \boldsymbol{\theta}) \text{ is nonsingular } \forall \boldsymbol{\theta} \in \Theta\}$. *We can write*

$$\int_\Theta \phi(\xi, \boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}) = \min_{\zeta\in\Xi^*} \sum_{\mathbf{x}\in\mathcal{X}} H_{\mathrm{AVE}}(\zeta, \mathbf{x})\,\xi(\mathbf{x}),$$

for any $\xi \in \Xi^*$, *where* $H_{\mathrm{AVE}}(\zeta, \mathbf{x}) = \int_\Theta H(\zeta, \mathbf{x}, \boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$ *and* $\Xi^* = \Xi_\Theta$ *for* D- *and* A-*optimality, and* $\Xi^* = \Xi$ *for the criteria of* E_k-*optimality.*

Proof. The design space $\mathcal{X}$ is assumed to be finite, hence the summation and the integration are interchangeable. From Theorems 2.1–2.3 we have for any $\zeta \in \Xi^*$ and for any $\boldsymbol{\theta} \in \Theta$ that

$$\phi(\xi, \boldsymbol{\theta}) \leq \sum_{\mathbf{x} \in \mathcal{X}} H(\zeta, \mathbf{x}, \boldsymbol{\theta}) \xi(\mathbf{x}). \tag{2.9}$$

We can write

$$\int_{\Theta} \phi(\xi, \boldsymbol{\theta}) \, d\pi(\boldsymbol{\theta}) \leq \sum_{\mathbf{x} \in \mathcal{X}} \left[\int_{\Theta} H(\zeta, \mathbf{x}, \boldsymbol{\theta}) \, d\pi(\boldsymbol{\theta}) \right] \xi(\mathbf{x}). \tag{2.10}$$

Since the inequality (2.10) holds for every $\zeta \in \Xi^*$, evidently:

$$\int_{\Theta} \phi(\xi, \boldsymbol{\theta}) \, d\pi(\boldsymbol{\theta}) \leq \min_{\zeta \in \Xi^*} \sum_{\mathbf{x} \in \mathcal{X}} \left[\int_{\Theta} H(\zeta, \mathbf{x}, \boldsymbol{\theta}) \, d\pi(\boldsymbol{\theta}) \right] \xi(\mathbf{x}). \tag{2.11}$$

The proofs of theorems 2.1–2.3 imply that for any $\boldsymbol{\theta} \in \Theta$ the equality in (2.9) is attained at $\zeta = \xi$, and so we obtain the equality in (2.10). This together with (2.11) proves the theorem. □

2.5 Bibliographic remarks on LP methods in experimental design

Since the cutting plane method based on LP is widely used in this thesis to obtain optimal designs, here we mention some other authors, who used LP for purposes of experimental design.

Gribik and Kortanek (1977) formulated an alternative to the Kiefer's equivalence theorem, so that they obtained conditions of the optimality of a design formulated as linear functions of this design, and hence they were able to apply LP methods. The difficulty with designs $\xi^{(i)}$ associated with singular information matrix (mentioned by Gribik and Kortanek, 1977, p. 245), which we overcome by the use of Remark 2.4, is solved in their paper differently requiring a solution of a nonlinear convex programming problem at each iteration. Moreover, their method requires the existence of a gradient of the considered criterion, so the E_k-criterion can not be handled. However, the advantage of the method of Gribik and Kortanek (1977) is the existence of a proof of convergence, obtained under some assumptions.

A possibility of an alternative approach to the results of Theorems 2.1–2.3 is indicated in Sect. 9.5.3 of Pronzato and Pázman (2013). The authors evaluated the subgradient of the considered criterion and then they applied the method of cutting planes of Kelley (1960) (see Sect. 1.4), which can be used for any concave criterion. Especially for the non-differentiable E_k criteria with $k \geq 2$, the approach presented in Theorem 2.3 seems to be more attractive than expressing the corresponding subgradients. As already mentioned in Sect. 2.1, the authors also used an alternative

formulation of the criteria of E-, $\mathbf{c}$-, and G-optimality, which allowed them to obtain optimal designs via LP.

Harman and Jurík (2008) consider only $\mathbf{c}$-optimal designs and they show that the famous Elfving's theorem can be interpreted as a formulation of an LP problem. This approach is quite different from ours.

2.6 Conclusions to this chapter

In Theorems 2.1–2.3 we demonstrate on the D-, A-, and E_k-optimality criteria that, by matrix manipulations, we can obtain in a straightforward way expressions which allow us to formulate the optimal design problem as an "infinite-dimensional" LP problem. The concavity and the positive homogeneity of these criteria are then evident. For a finite design space, the application of the relaxation method then leads not only to a simple algorithm for optimum design, but also to stopping rules different from the standard Kiefer's equivalence theorem. Remark 2.4 is also important as it justifies regularization steps in the algorithm.

Although the literature devoted to the optimal experimental design offers a lot of methods and algorithms to obtain optimal designs which maximize standard criteria like D- or A-optimality (see e.g. Chap. 9 of Pronzato and Pázman, 2013 for a review of some methods), the methods described in this chapter may be very useful to optimize some more complicated criteria. For instance to compute the criterion robust design of Harman (2004) as presented in Sect. 2.3.1 or the optimization with respect to one criterion under a bound constraint on the value of another criterion, see Sect. 2.3.2. For more detailed results and examples we refer the reader to our paper Burclová and Pázman (2016a).

We used the method of cutting planes also to compute the CVaR optimal designs in nonlinear models, see Chap. 3, and to maximize the extended optimality criteria in generalized regression models, where we used the same algorithm as suggested by Pázman and Pronzato (2014), see Chap. 4.

Chapter 3

Experimental design based on some ideas from risk theory

A special attention in this chapter is paid to the criterion based on the "Conditional Value at Risk" (CVaR, see Sect. 3.3.1). Conditional Value at Risk, also called sometimes "super-quantile", and "Value at Risk" (VaR, see Sect. 3.3.1), or simply "quantile", are the terms commonly used in the risk theory (finance and actuarial sciences), but they can be applied also in the industry and engineering, see e.g. Guerra (2016).

The Value at Risk and Conditional Value at Risk are in the literature (e.g. Pflug, 2000; Rockafellar and Uryasev, 2000, 2002) considered as functions of a random variable Y which measures the random loss associated with an investor decision, and are supposed to be minimized. Unless indicated otherwise, when we will refer to these papers we will use the notation common in experimental design, i.e. the random loss corresponds to the criterion $\phi(\xi, \boldsymbol{\theta})$ and is to be maximized.

Generally, the local, AVE, and maximin criteria from Def. 1.3 have their pros and cons, for a recent exposition we refer to Sects. 8.1, 8.2, and 8.4 in Pronzato and Pázman (2013). Due to some undesirable properties of above-mentioned criteria, Pázman and Pronzato (2007) suggested to use **the quantile criterion** for a prescribed value $\alpha \in [0, 1]$

$$\Phi_\alpha^Q(\xi) = \max\{t \in \mathbb{R} : Pr[\phi(\xi, \boldsymbol{\theta}) \geq t] \geq 1 - \alpha\}. \tag{3.1}$$

The quantile criterion is closely related to VaR and has many nice properties, see e.g. Pázman and Pronzato (2007) or Sect. 8.4 in Pronzato and Pázman (2013). But, generally, the quantile criterion is neither concave nor convex, which makes the computation of an optimal design very complicated.

Valenzuela et al. (2015) introduced another (convex) criterion based on CVaR, which has very similar properties as the quantile criterion. However, in the computations Valenzuela et al. (2015) used a formula developed by Rockafellar and Uryasev

(2000), which can be used (in their settings) only when $\phi(\xi, \boldsymbol{\theta})$ is a continuous random variable. It is obvious that the countability or finiteness of the parametric space implies that $\phi(\xi, \boldsymbol{\theta})$ is a discrete random variable. Rockafellar and Uryasev (2002) extended the results of Rockafellar and Uryasev (2000) also for noncontinuous distributions using different definition of CVaR. So in the thesis we rely on papers Rockafellar and Uryasev (2002) and Pflug (2000), which help us to interpret and analyse the CVaR criterion for discrete $\phi(\xi, \boldsymbol{\theta})$.

Further technical problem is our preference for the concave definition of the CVaR-criterion to be consistent with the other parts of this thesis. This is attained by a simple reformulation of CVaR.

After defining some basic terms in Sect. 3.1, in Sect. 3.2 we define the criterion related to CVaR following Pflug (2000). The criterion is concave when $\phi(\xi, \boldsymbol{\theta})$ is concave in ξ regardless of continuity of its cdf. The analogies of statements in this section can be found in the literature devoted to the risk theory (e.g. Rockafellar and Uryasev, 2000; Pflug, 2000; Rockafellar and Uryasev, 2002), but here we formulate also some proofs for readers convenience. Section 3.3 sheds some light on the interpretation of our quite abstract CVaR-criterion and is inspired by the paper Rockafellar and Uryasev (2002). In Sect. 3.4 we show the relation of the CVaR-criterion to the AVE, maximin and quantile criteria and we formulate the equivalence theorem. In Sect. 3.5 we implement the cutting plane method to maximize the CVaR-criterion. The algorithm is tested on the examples in Sect. 3.6. Section 3.7 concludes this chapter.

3.1 Basic definitions

Let $\pi(\boldsymbol{\theta})$ be the prior distribution of unknown parameter $\boldsymbol{\theta}$ on the parametric space Θ. In this sense, $\boldsymbol{\theta}$ is a random variable or a random vector taking values from the space Θ. Consequently, the information matrix $M(\xi, \boldsymbol{\theta})$ associated with the design ξ and the criterion value $\phi(\xi, \boldsymbol{\theta})$ are random.

For purposes of this chapter, let us define the basic terms.

Definition 3.1. The cumulative distribution function (cdf) of a random variable Y is defined as

$$F_Y(y) = Pr(Y \leq y) \in [0, 1].$$

For $\alpha \in [0, 1]$ we define the α-quantile (sometimes referred as α-left-quantile) of Y as

$$q_Y(\alpha) = \inf\{y \in \mathbb{R} : F_Y(y) \geq \alpha\} \tag{3.2}$$

and the α-right-quantile of Y as

$$\begin{aligned} Rq_Y(\alpha) &= \sup\{y \in \mathbb{R} : Pr(Y \geq y) \geq 1-\alpha\} \\ &= \sup\{y \in \mathbb{R} : Pr(Y < y) \leq \alpha\}. \end{aligned} \tag{3.3}$$

For the term right-quantile and for more properties of quantiles we refer to Hosseini (2010, 2009).

Note that for $\alpha \in (0,1)$ one may write minimum instead of infimum in (3.2) because the set $\{y \in \mathbb{R} : F_Y(y) \geq \alpha\}$ is the interval bounded from below including the endpoint, since $F(y)$ is non-decreasing and right-continuous function. Similarly, the set $\{y \in \mathbb{R} : Pr(Y \geq y) \geq 1-\alpha\} = \{y \in \mathbb{R} : Pr(Y < y) \leq \alpha\}$ is the interval bounded from above including the endpoint because the function $Pr(Y < y)$ is non-decreasing and left-continuous and we can write maximum instead of supremum in (3.3).

Definition 3.2. A random variable Y is:

- **discrete** when it is taking finitely or countably many values and P_Y is its probability mass function (pmf) $P_Y(B) = Pr\left[Y^{-1}(B)\right]$ for any Borel set B,
- **continuous** when it is taking uncountably many values and its cdf $F_Y(y)$ satisfies $F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt$, where the function $f_Y(\cdot)$ is the probability density function (pdf) of the random variable Y.

Note that when the random vector $\boldsymbol{\theta}$ has a discrete distribution, then also the criterion value $\phi(\xi, \boldsymbol{\theta})$ is distributed discretely.

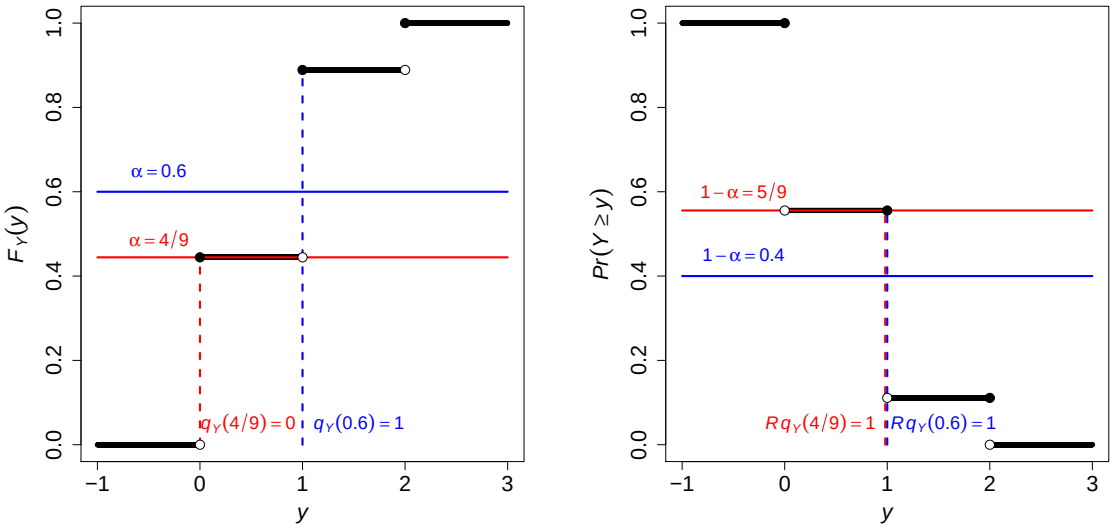
When $F_Y(y)$ takes the value $\alpha \in (0,1]$ if and only if y is from $\mathcal{I}$, some interval in $\mathbb{R}$, then the α-left-quantile corresponds to the left endpoint of $\mathcal{I}$ and the α-right-quantile corresponds to the right endpoint of $\mathcal{I}$. This property is illustrated in the next example.

Example 3.3. Consider a random variable $Y \sim Bin(2, 1/3)$ taking values $0, 1, 2$. Its cdf (continuous from the right) and the function $Pr(Y \geq y)$ (continuous from the left) are displayed in Figure 3.1. For $\alpha = 0.6$ we obtain $q_Y(\alpha) = Rq_Y(\alpha) = 1$, but for $\alpha = 4/9 = F_Y(0)$ we have $q_Y(\alpha) = 0$ and $Rq_Y(\alpha) = 1$. △

One obtains for any random variable Y that

$$\begin{aligned} -q_{-Y}(\alpha) = -\inf\{y : Pr(-Y \leq y) \geq \alpha\} = \sup\{-y : Pr(Y \geq -y) \geq \alpha\} = \\ = \sup\{y : Pr(Y \geq y) \geq \alpha\} = Rq_Y(1-\alpha). \end{aligned} \tag{3.4}$$

Since the functions $y \mapsto Pr(Y < y)$ and $y \mapsto F_Y(y) = Pr(Y \leq y)$ are non-decreasing functions, the first is continuous from the left, the second from the right, the set $\mathcal{I} = \{y : Pr(Y < y) \leq \alpha\}$ forms a right-closed interval (half-line $(-\infty, Rq_Y(\alpha)]$), and



(a) Cumulative distribution function

(b) Decumulative distribution function

Figure 3.1: Example 3.3: The relation of $q_Y(\alpha)$ to $Rq_Y(\alpha)$ when $Y \sim Bin(2, 1/3)$.

the set $\mathcal{I}' = \{y : Pr(Y \leq y) > \alpha\}$ forms a left-bounded interval (half-line $[Rq_Y(\alpha), \infty)$ or $(Rq_Y(\alpha), \infty)$) with the same endpoint $Rq_Y(\alpha) \in \mathcal{I}$. One sees that

$$Rq_Y(\alpha) = \sup_{y \in \mathcal{I}} y = \inf_{y \in \mathcal{I}'} y = \inf\{y \in \mathbb{R} : F_Y(y) > \alpha\}. \tag{3.5}$$

This corresponds to the definition of right-quantile in Hosseini (2010) and Rockafellar and Uryasev (2002). It follows from (3.2) and (3.5) that

$$q_Y(\alpha) = \inf\{y : F_Y(y) \geq \alpha\} \leq \inf\{y : F_Y(y) > \alpha\} = Rq_Y(\alpha). \tag{3.6}$$

3.2 CVaR-criterion for experimental design purposes

Definition 3.4. We define the CVaR optimality criterion for $\alpha \in (0, 1]$ as

$$\Phi_\alpha(\xi) \equiv \max_{c \in \mathbb{R}} \left\{ c + \frac{1}{\alpha} E[\min\{0, \phi(\xi, \boldsymbol{\theta}) - c\}] \right\}. \tag{3.7}$$

We used the analogous definition of CVaR as Pflug (2000), but the formula was originally derived in Rockafellar and Uryasev (2000). The relation of the criterion to the theory of risk is pointed out in Sect. 3.3. Below we show some of its useful properties.

3.2.1 Concavity of the CVaR-criterion

For the sake of completeness we give the proof of the concavity of the CVaR-criterion Φ_α. Our proofs are based on the proofs from the theory risk, where the convexity of similar functions is proved.

We introduce the following notation:

$$
\begin{aligned}
w_\alpha(\xi, c) &\equiv c + \frac{1}{\alpha} E\left[\min\{0, \phi(\xi, \boldsymbol{\theta}) - c\}\right], \\
w_{\alpha,Y}(c) &\equiv c + \frac{1}{\alpha} E\left[\min\{0, Y - c\}\right],
\end{aligned}
\tag{3.8}
$$

where Y is any random variable. Then one can write

$$
\Phi_\alpha(\xi) = \max_{c \in \mathbb{R}} w_\alpha(\xi, c) = \max_{c \in \mathbb{R}} w_{\alpha, \phi(\xi, \boldsymbol{\theta})}(c). \tag{3.9}
$$

Lemma 3.5. *Let the criterion $\phi(\xi, \boldsymbol{\theta})$ be concave in ξ for any $\boldsymbol{\theta} \in \Theta$ and let Y be a random variable. Then for a given $\alpha \in (0, 1]$*

a) the function $w_{\alpha,Y}(c) \equiv c + \frac{1}{\alpha} E\left[\min\{0, Y - c\}\right]$ is concave in c for any Y, i.e. for any $\lambda \in (0,1)$ and for any $c_1, c_2 \in \mathbb{R}$ we have $w_{\alpha,Y}\left[\lambda c_1 + (1-\lambda) c_2\right] \geq \lambda w_{\alpha,Y}(c_1) + (1-\lambda) w_{\alpha,Y}(c_2)$,

b) the function $w_\alpha(\xi, c) \equiv c + \frac{1}{\alpha} E\left[\min\{0, \phi(\xi, \boldsymbol{\theta}) - c\}\right]$ is concave in (ξ, c), i.e. for any $\lambda \in (0,1)$, $\xi_1, \xi_2 \in \Xi$ and $c_1, c_2 \in \mathbb{R}$ we have $w_\alpha\left[\lambda \xi_1 + (1-\lambda)\xi_2, \lambda c_1 + (1-\lambda) c_2\right]$ $\geq \lambda w_\alpha(\xi_1, c_1) + (1-\lambda) w_\alpha(\xi_2, c_2)$.

Proof. Notice that we postponed analogically to proofs of the convexity of functions similar to $w_{\alpha,Y}(\xi, c)$ resp. $w_\alpha(\xi, c)$ which are given in Proposition 2(iv) of Pflug (2000), Theorem 2 of Rockafellar and Uryasev (2000), Theorem 10 and Corollary 11 of Rockafellar and Uryasev (2002).

a) The function $c \mapsto \min\{0, z - c\}$ is for any z concave in c, since the minimum of two concave functions is concave. Hence for any $\lambda \in (0,1)$

$$
\begin{aligned}
& w_{\alpha,Y}[\lambda c_1 + (1-\lambda) c_2] \\
&= \lambda c_1 + (1-\lambda) c_2 + \frac{1}{\alpha} E\left[\min\{0, Y - \lambda c_1 - (1-\lambda) c_2\}\right] \\
&\geq \lambda c_1 + (1-\lambda) c_2 + \frac{1}{\alpha} E\left[\lambda \min\{0, Y - c_1\} + (1-\lambda) \min\{0, Y - c_2\}\right] \\
&= \lambda w_{\alpha,Y}(c_1) + (1-\lambda) w_{\alpha,Y}(c_2).
\end{aligned}
$$

b) The functions $(\xi, c) \mapsto 0$ and $(\xi, c) \mapsto \phi(\xi, \boldsymbol{\theta}) - c$ are concave in (ξ, c) and hence $(\xi, c) \mapsto \min\{0, \phi(\xi, \boldsymbol{\theta}) - c\}$ is concave. Proceeding similarly as in the case a) one obtains that the function $w_\alpha(\xi, c)$ is concave in (ξ, c). □

Theorem 3.6. *When $\phi(\xi, \boldsymbol{\theta})$ is concave in ξ for any $\boldsymbol{\theta} \in \Theta$, then for given $\alpha \in (0, 1]$ the criterion $\Phi_\alpha(\xi)$ is concave in ξ.*

Proof. The proof is according to Proposition 2(iv) in Pflug (2000), who proved the convexity of a function similar to $\Phi_\alpha(\xi)$. Using the notation from (3.8) we have $\Phi_\alpha(\xi) = \max_{c\in\mathbb{R}} w_\alpha(\xi, c)$. Let us denote

$$c_1 \in \arg\max_{c\in\mathbb{R}} w_\alpha(\xi_1, c),$$
$$c_2 \in \arg\max_{c\in\mathbb{R}} w_\alpha(\xi_2, c),$$

where ξ_1 and ξ_2 are from Ξ. Then for every $\lambda \in (0,1)$

$$\Phi_\alpha[\lambda\xi_1 + (1-\lambda)\xi_2] = \max_{c\in\mathbb{R}} w_\alpha[\lambda\xi_1 + (1-\lambda)\xi_2, c] \geq w_\alpha[\lambda\xi_1 + (1-\lambda)\xi_2, \lambda c_1 + (1-\lambda)c_2],$$

and using the concavity of $w_\alpha(\cdot,\cdot)$ (see Lemma 3.5b) we have

$$\Phi_\alpha[\lambda\xi_1 + (1-\lambda)\xi_2] \geq \lambda w_\alpha(\xi_1, c_1) + (1-\lambda) w_\alpha(\xi_2, c_2) = \lambda\Phi_\alpha(\xi_1) + (1-\lambda)\Phi_\alpha(\xi_2).$$

□

3.2.2 Conditions for finiteness of the CVaR-criterion

Lemma 3.7. *For any given $\alpha \in (0,1]$ the function $w_{\alpha,Y}(c)$ is bounded from above by $E(Y)$. Moreover, $\max_{c\in\mathbb{R}} w_{\alpha,Y}(c) \geq -\frac{1}{\alpha}E(|Y|)$.*

Proof. The function $z \mapsto \min\{0, z-c\}$ is concave and hence it follows from the Jensen's inequality (see e.g. Feller, 1971, Eq. 8.6 or Billingsley, 1995, Eq. 5.33) that if $E(Y)$ exists, then

$$\begin{aligned} w_{\alpha,Y}(c) &= c + \frac{1}{\alpha}E(\min\{0, Y-c\}) \\ &\leq c + \frac{1}{\alpha}\min\{0, E(Y)-c\} = \begin{cases} c & \text{if } E(Y) \geq c, \\ c + \frac{1}{\alpha}[E(Y)-c] & \text{if } E(Y) < c, \end{cases} \\ &\leq \begin{cases} \max\limits_{c: E(Y)\geq c} c & \text{if } E(Y) \geq c, \\ \max\limits_{c: E(Y)<c} c + \frac{1}{\alpha}[E(Y)-c] & \text{if } E(Y) < c, \end{cases} = E(Y). \end{aligned}$$

On the other hand,

$$\begin{aligned} \max_{c\in\mathbb{R}} w_{\alpha,Y}(c) \geq w_{\alpha,Y}(0) &= \frac{1}{\alpha}E(\min\{0, Y\}) = -\frac{1}{\alpha}E(|Y| \mid Y<0)\,Pr(Y<0) \geq \\ &\geq -\frac{1}{\alpha}E(|Y| \mid Y<0)\,Pr(Y<0) - \frac{1}{\alpha}E(|Y| \mid Y\geq 0)\,Pr(Y\geq 0) = -\frac{1}{\alpha}E(|Y|). \end{aligned}$$

□

According to the previous lemma, the finiteness of the criterion $\Phi_\alpha(\xi) = \max_{c\in\mathbb{R}} w_{\alpha,\phi(\xi,\boldsymbol{\theta})}(c)$ is ensured when the random variable $|Y| = |\phi(\xi, \boldsymbol{\theta})|$ has finite expectation:

Consequence 3.8. When $E[|\phi(\xi,\boldsymbol{\theta})|] < \infty$ for $\xi \in \Xi$, then $|\Phi_\alpha(\xi)| < \infty$ for any $\alpha \in (0,1]$.

3.2.3 Continuity of the function $w_\alpha(\xi,c)$ in the formulation of the CVaR-criterion in (3.9)

Lemma 3.9. *Consider the random variable Y with $E(|Y|) < \infty$. Then the function $w_{\alpha,Y}(c) = c + \frac{1}{\alpha}E(\min\{0, Y-c\})$ is continuous in c on every closed bounded subset of $\mathbb{R}$. Moreover, when $\mathcal{X}$ is finite and $\phi(\xi,\boldsymbol{\theta})$ is concave function of ξ with $E(|\phi(\xi,\boldsymbol{\theta})|) < \infty$ for every $\xi \in \Xi$, then the function $w_\alpha(\xi,c) = c + \frac{1}{\alpha}E(\min\{0, \phi(\xi,\boldsymbol{\theta})-c\})$ is continuous in $(\xi,c) \in \Xi \times C$, where C is an arbitrary closed bounded subset of $\mathbb{R}$.*

Proof. The function $w_{\alpha,Y}(c)$ is concave (Lemma 3.5a) and, moreover, it is finite for any $c \in \mathbb{R}$ because $E(|Y|) \leq \infty$ (Lemma 3.7). As a consequence of Theorem 35.1 in Rockafellar (1970), the concavity and finiteness of $w_{\alpha,Y}(c)$ imply that it is also continuous in c for any $\alpha \in (0,1]$. The continuity of $w_\alpha(\xi,c)$ is proved in the same way using Lemma 3.5b and Consequence 3.8. □

Note that the continuity in c is also mentioned in Theorem 10 of Rockafellar and Uryasev (2002).

Example 3.3. (continued) The continuity of function $w_{\alpha,Y}(c)$ for $Y \sim Bin(2,1/3)$ is illustrated in Figure 3.2a. One sees that in the case $\alpha = 0.6$, the maximum of $w_{\alpha,Y}(c)$ is reached at $c = 1$, on the other hand, for $\alpha = 4/9 = F_Y(0)$ is the maximum attained at any $c \in [0,1]$ (the location of the maximum is justified in Theorem 3.10).

3.2.4 Optimal c in the definition of CVaR-criterion Φ_α in (3.7)

Theorem 3.10. *Let Y be any random variable. Then $[q_Y(\alpha), Rq_Y(\alpha)] = \arg\max_{c\in\mathbb{R}} w_{\alpha,Y}(c)$, where $w_{\alpha,Y}(c) = c + \frac{1}{\alpha}E(\min\{0, Y-c\})$ and $\alpha \in (0,1]$ is given.*

Proof. In Theorem 10 of Rockafellar and Uryasev (2002) (see also Proposition 1 of Pflug (2000)) is proved that

$$[q_{-Y}(1-\alpha), Rq_{-Y}(1-\alpha)] = \arg\min_{c\in\mathbb{R}} c + \frac{1}{\alpha}E[\max\{0, -Y-c\}] = \arg\min_{c\in\mathbb{R}} -w_{\alpha,Y}(-c).$$

If $c^\star \in \arg\max_{c\in\mathbb{R}} w_{\alpha,Y}(c)$, then $c^* \in \arg\min_{c\in\mathbb{R}} -w_{\alpha,Y}(c)$ and hence

$$-c^\star \in \arg\min_{c\in\mathbb{R}} -w_{\alpha,Y}(-c) = [q_{-Y}(1-\alpha), Rq_{-Y}(1-\alpha)].$$

It follows that $c^\star \in [-Rq_{-Y}(1-\alpha), -q_{-Y}(1-\alpha)] = [q_Y(\alpha), Rq_Y(\alpha)]$, where we used (3.4). □

Notice that when $\alpha = 1$, then $Rq_Y(\alpha) = Rq_Y(1) = \infty$ and hence the maximum of $w_{1,Y}(c)$ is reached in the half-open interval $[q_Y(1), Rq_Y(1))$. If, moreover, $q_Y(1) = \infty$ for some random variable Y (e.g. $Y \sim \mathcal{N}(0,1)$), the maximum of $w_{1,Y}(c)$ is reached for $c \to \infty$ (and supremum is more appropriate). But, according to Lemma 3.7, the value $\sup_c w_{1,Y}(c)$ stays finite when $E(|Y|) < \infty$ and, in addition, $\sup_c w_{1,Y}(c) = E(Y)$ (see also Theorem 3.19a).

Consequence 3.11. Alternatively, the CVaR-criterion (3.7) can be defined as follows

$$\Phi_\alpha(\xi) = c^\star + \frac{1}{\alpha} E[\min\{0, \phi(\xi, \boldsymbol{\theta}) - c^\star\}],$$

where $c^\star$ is any point from $\left[q_{\phi(\xi,\boldsymbol{\theta})}(\alpha), Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]$.

In the proofs of this chapter we will often use the Law of total expectation (see e.g. Billingsley, 1995, Eq. 34.6): $E(\min\{Y-c, 0\}) = E(Y - c \mid Y \le c) Pr(Y \le c) = E(Y \mid Y \le c) Pr(Y \le c) - cPr(Y \le c)$ (or analogically with the strict inequalities), and hence (supposing that the probabilities in the denominators are positive)

$$E(Y \mid Y \le c) = \frac{E(\min\{Y-c, 0\})}{Pr(Y \le c)} + c \text{ and } E(Y \mid Y < c) = \frac{E(\min\{Y-c, 0\})}{Pr(Y < c)} + c. \tag{3.10}$$

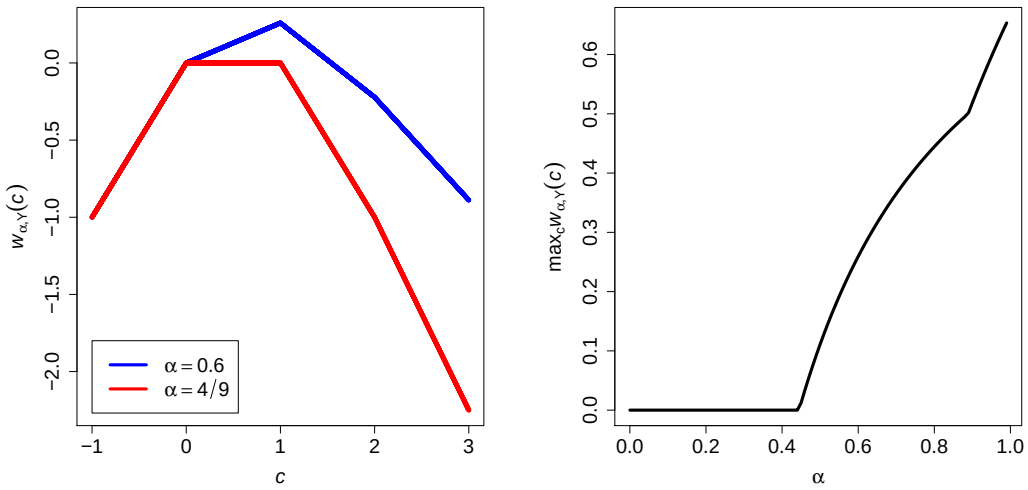
Further, one obtains for $c^\star \in \left[q_{\phi(\xi,\boldsymbol{\theta})}(\alpha), Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]$

$$\begin{aligned} \Phi(\xi, \boldsymbol{\theta}) &= c^\star \left\{1 - \frac{Pr[\phi(\xi, \boldsymbol{\theta}) \le c^\star]}{\alpha}\right\} + E[\phi(\xi, \boldsymbol{\theta}) \mid \phi(\xi, \boldsymbol{\theta}) \le c^\star] \frac{Pr[\phi(\xi, \boldsymbol{\theta}) \le c^\star]}{\alpha} \\ &= c^\star \left\{1 - \frac{Pr[\phi(\xi, \boldsymbol{\theta}) < c^\star]}{\alpha}\right\} + E[\phi(\xi, \boldsymbol{\theta}) \mid \phi(\xi, \boldsymbol{\theta}) < c^\star] \frac{Pr[\phi(\xi, \boldsymbol{\theta}) < c^\star]}{\alpha}. \end{aligned} \tag{3.11}$$

3.2.5 The dependence of the CVaR-criterion Φ_α on the probability level α

Lemma 3.12. *The function* $\max_{c \in \mathbb{R}} w_{\alpha,Y}(c)$ *is non-decreasing in* $\alpha \in (0,1]$ *for any random variable* Y.

Proof. Let $1 \ge \alpha > \beta > 0$. One sees from non-decreasing tendency of F_Y, from (3.5) and from (3.6) that $q_Y(\beta) \le Rq_Y(\beta) = \inf\{y : F_Y(y) > \beta\} \le \inf\{y : F_Y(y) \ge \alpha\} = q_Y(\alpha) \le Rq_Y(\alpha)$, and hence, following Theorem 3.10, $c_\alpha \ge c_\beta$, for arbitrarily chosen $c_\alpha \in \arg\max_{c \in \mathbb{R}} w_{\alpha,Y}(c)$, $c_\beta \in \arg\max_{c \in \mathbb{R}} w_{\beta,Y}(c)$. Then using this we obtain that



(a) The continuity of function $w_{\alpha,Y}(c)$ for $Y \sim Bin(2,1/3)$ and $c \in [-1,3]$.

(b) The non-decreasing tendency of $\max_{c\in\mathbb{R}} w_{\alpha,Y}(c)$ in α for $Y \sim Bin(2,1/3)$.

Figure 3.2: Example 3.3: The illustration of Lemma 3.9 and Lemma 3.12.

the difference

$$\begin{aligned}
\max_{c\in\mathbb{R}} w_{\alpha,Y}(c) - \max_{c\in\mathbb{R}} w_{\beta,Y}(c) &= c_\alpha - c_\beta + \frac{1}{\alpha}E\left[\min\{0, Y-c_\alpha\}\right] - \frac{1}{\beta}E\left[\min\left\{0, Y-c_\beta\right\}\right] \\
&\geq c_\alpha - c_\beta + \frac{1}{\alpha}E\left[\min\{0, Y-c_\alpha\} + \max\left\{0, c_\beta - Y\right\}\right] \\
&\geq c_\alpha - c_\beta + \frac{1}{\alpha}0Pr\left[Y \geq c_\alpha\right] \\
&\quad + \frac{1}{\alpha}E\left[Y - c_\alpha + c_\beta - Y \mid Y < c_\alpha\right] Pr\left[Y < c_\alpha\right] \\
&\geq c_\alpha - c_\beta - \frac{1}{\alpha}\left(c_\alpha - c_\beta\right) Pr\left[Y < Rq_Y(\alpha)\right] \\
&\geq c_\alpha - c_\beta - \frac{1}{\alpha}\left(c_\alpha - c_\beta\right)\alpha = 0
\end{aligned}$$

is nonnegative, and the criterion $\Phi_\alpha(\xi)$ is non-decreasing function of parameter α. □

Example 3.3. (continued) The non-decreasing tendency of function $\alpha \mapsto w_{\alpha,Y}(c)$ for $Y \sim Bin(2,1/3)$ is illustrated in Figure 3.2b.

Consequence 3.13. The criterion $\Phi_\alpha(\xi)$ is non-decreasing in $\alpha \in (0,1]$ for any given ξ.

3.3 Two easily interpreted criteria as lower and upper bound of the CVaR-criterion

If $\phi(\xi, \boldsymbol{\theta})$ is a discrete random variable, one could restrict his attention only to "appropriate alphas", such that for some value $t_{\alpha,\xi} \in \mathbb{R}$ $F_{\phi(\xi,\boldsymbol{\theta})}\left(t_{\alpha,\xi}\right) = \alpha$. Then

$F_{\phi(\xi,\boldsymbol{\theta})}\left[q_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]=\alpha$ and according to (3.11)

$$\begin{aligned}\Phi_\alpha(\xi) &= q_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\left\{1-\frac{F_{\phi(\xi,\boldsymbol{\theta})}\left[q_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]}{\alpha}\right\}\\ &\quad + E\left[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) \le q_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]\frac{F_{\phi(\xi,\boldsymbol{\theta})}\left[q_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]}{\alpha}\\ &= E\left[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) \le q_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right].\end{aligned}$$

The CVaR-criterion can be then interpreted straightforwardly. But when applying the criterion $\Phi_\alpha(\xi)$ in experimental design, the set of "appropriate alphas" may change with different designs ξ. So, in design problems, the restriction to "appropriate alphas" is not possible and for these reasons we introduce further optimality criteria based on the risk theory, which, as we know, were not analysed for the purposes of experimental design, but they may help us to interpret and understand the criterion given in (3.7) in terms of conditional expectation.

3.3.1 Formulation of some optimality criteria based on the risk theory

Let Y be a random variable (denoting for instance the loss associated with an investment possibility). In papers related to the risk theory, the following variables were considered and analysed:

$$\begin{aligned}&\mathrm{VaR}_\beta(Y) \equiv \min\{y : F_Y(y) \ge \beta\} = q_Y(\beta) && \beta\in[0,1] \text{ (Value at Risk)},\\ &\mathrm{CVaR}_\beta(Y) \equiv \min_{c\in\mathbb{R}}\left[c+\frac{1}{1-\beta}E(\max\{0,Y-c\})\right] && \beta\in[0,1) \text{ (Conditional Value at Risk)},\\ &\mathrm{CVaR}^+_\beta(Y) \equiv E\left[Y \mid Y > \mathrm{VaR}_\beta(Y)\right] && \beta\in[0,1] \text{ (Upper CVaR)},\\ &\mathrm{CVaR}^-_\beta(Y) \equiv E\left[Y \mid Y \ge \mathrm{VaR}_\beta(Y)\right] && \beta\in[0,1] \text{ (Lower CVaR)},\end{aligned}$$

see Rockafellar and Uryasev (2000, 2002); Pflug (2000) for the definition of VaR, Pflug (2000) for *this* definition of CVaR and Rockafellar and Uryasev (2002) for CVaR^+ and CVaR^-.

Since Y describes the possible loss, the aim of the risk theory is to minimize VaR, CVaR, CVaR^+, or CVaR^- by proper choice of the investment strategy. For the purposes of experimental design, we have to take instead of Y the expression $-\phi(\xi,\boldsymbol{\theta})$ and maximize the negatives of $\mathrm{VaR}_\beta[-\phi(\xi,\boldsymbol{\theta})]$, $\mathrm{CVaR}_\beta[-\phi(\xi,\boldsymbol{\theta})]$, $\mathrm{CVaR}^+_\beta[-\phi(\xi,\boldsymbol{\theta})]$ and $\mathrm{CVaR}^-_\beta[-\phi(\xi,\boldsymbol{\theta})]$.

One can obtain

$$\begin{aligned}
-\mathrm{CVaR}_\beta\left[-\phi\left(\xi,\boldsymbol{\theta}\right)\right] = - & \min_{c\in\mathbb{R}}\left\{c+\frac{1}{1-\beta}E\left[\max\left\{0,-\phi\left(\xi,\boldsymbol{\theta}\right)-c\right\}\right]\right\} \\
= & \max_{c\in\mathbb{R}}\left\{-c-\frac{1}{1-\beta}E\left[\max\left\{0,-\phi\left(\xi,\boldsymbol{\theta}\right)-c\right\}\right]\right\} \\
= & \max_{c\in\mathbb{R}}\left\{-c+\frac{1}{1-\beta}E\left[\min\left\{0,\phi\left(\xi,\boldsymbol{\theta}\right)+c\right\}\right]\right\} \\
= & \max_{c\in\mathbb{R}}\left\{c+\frac{1}{1-\beta}E\left[\min\left\{0,\phi\left(\xi,\boldsymbol{\theta}\right)-c\right\}\right]\right\} = \Phi_{1-\beta}\left(\xi\right).
\end{aligned}$$

So for $\alpha = 1-\beta$ the last expression equals to the CVaR-criterion $\Phi_\alpha\left(\xi\right)$ defined in (3.7). In the risk theory, the usual choices for β are 0.9, 0.95 or 0.99 (c.f. e.g. Rockafellar and Uryasev, 2000), but here we are not limited only to these special choices for α and we will consider different probability levels, e.g. $\alpha = 0.5$ to obtain an alternative for median criterion, which can be interesting in the applications.

The quantile criterion from (3.1) can be interpreted in the terms of VaR

$$-\mathrm{VaR}_\beta\left[-\phi\left(\xi,\boldsymbol{\theta}\right)\right] = -q_{-\phi\left(\xi,\boldsymbol{\theta}\right)}\left(\beta\right) = Rq_{\phi\left(\xi,\boldsymbol{\theta}\right)}\left(1-\beta\right) = \Phi^Q_{1-\beta}\left(\xi\right)$$

(see Eq. (3.4)). Similarly,

$$\begin{aligned}
-\mathrm{CVaR}^+_\beta\left[-\phi\left(\xi,\boldsymbol{\theta}\right)\right] &= -E\left[-\phi\left(\xi,\boldsymbol{\theta}\right) \mid -\phi\left(\xi,\boldsymbol{\theta}\right) > \mathrm{VaR}_\beta\left(-\phi\left(\xi,\boldsymbol{\theta}\right)\right)\right] \\
&= E\left[\phi\left(\xi,\boldsymbol{\theta}\right) \mid \phi\left(\xi,\boldsymbol{\theta}\right) < -\mathrm{VaR}_\beta\left(-\phi\left(\xi,\boldsymbol{\theta}\right)\right)\right] \\
&= E\left[\phi\left(\xi,\boldsymbol{\theta}\right) \mid \phi\left(\xi,\boldsymbol{\theta}\right) < \Phi^Q_{1-\beta}\left(\xi\right)\right],
\end{aligned}$$

and we define

$$\Phi^-_{1-\beta}\left(\xi\right) = E\left[\phi\left(\xi,\boldsymbol{\theta}\right) \mid \phi\left(\xi,\boldsymbol{\theta}\right) < \Phi^Q_{1-\beta}\left(\xi\right)\right]$$

and

$$\Phi^+_{1-\beta}\left(\xi\right) = E\left[\phi\left(\xi,\boldsymbol{\theta}\right) \mid \phi\left(\xi,\boldsymbol{\theta}\right) \leq \Phi^Q_{1-\beta}\left(\xi\right)\right]$$

(the notation is summarized below in Def. 3.14). Notice that $\Phi^-_{1-\beta}\left(\xi\right)$ is not well-defined for such β that $Pr\left[\phi\left(\xi,\boldsymbol{\theta}\right) < \Phi^Q_{1-\beta}\left(\xi\right)\right] = 0$, i.e. when $\Phi^Q_{1-\beta}\left(\xi\right)$ corresponds to the smallest possible value of the random variable $\phi\left(\xi,\boldsymbol{\theta}\right)$.

We put $\alpha = 1-\beta$ and from now we will focus on criteria which are supposed to be maximized.

Definition 3.14. Let $\phi\left(\xi,\boldsymbol{\theta}\right)$ be the optimality criterion dependent on unknown parameter $\boldsymbol{\theta}$. Then for $\alpha \in (0,1]$ we define the following optimality criteria:

$$\begin{aligned}
\Phi^Q_\alpha\left(\xi\right) &= \max\left\{t \in \mathbb{R} : Pr\left[\phi\left(\xi,\boldsymbol{\theta}\right) \geq t\right] \geq 1-\alpha\right\} = Rq_{\phi\left(\xi,\boldsymbol{\theta}\right)}\left(\alpha\right), \\
\Phi_\alpha\left(\xi\right) &= \max_{c\in\mathbb{R}}\left\{c+\frac{1}{\alpha}E\left[\min\left\{0,\phi\left(\xi,\boldsymbol{\theta}\right)-c\right\}\right]\right\}, \\
\Phi^+_\alpha\left(\xi\right) &= E\left[\phi\left(\xi,\boldsymbol{\theta}\right) \mid \phi\left(\xi,\boldsymbol{\theta}\right) \leq \Phi^Q_\alpha\left(\xi\right)\right].
\end{aligned}$$

For α such that $Pr\left[\phi(\xi,\boldsymbol{\theta}) < \Phi_\alpha^Q(\xi)\right] > 0$ we also define the criterion

$$\Phi_\alpha^-(\xi) = E\left[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) < \Phi_\alpha^Q(\xi)\right].$$

Note that we have already defined the criterion $\Phi_\alpha(\xi)$ in Def. 3.4 and the criterion $\Phi_\alpha^Q(\xi)$ in (3.1).

3.3.2 The comparison of the criteria based on the risk theory

According to (3.6) and Consequence 3.11, it is easy to see that for any $\xi \in \Xi$ and for arbitrarily chosen $\alpha \in (0,1]$

$$\Phi_\alpha^Q(\xi) = Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha) \geq q_{\phi(\xi,\boldsymbol{\theta})}(\alpha) \geq \Phi_\alpha(\xi). \tag{3.12}$$

Further bounds on $\Phi_\alpha(\xi)$ are provided in the next theorem, see also Proposition 5 of Rockafellar and Uryasev (2002) for the analogous statement.

Theorem 3.15. *Let the level $\alpha \in (0,1]$ be given. When $\Phi_\alpha^-(\xi)$ is well-defined, i.e. when $Pr\left[\phi(\xi,\boldsymbol{\theta}) < \Phi_\alpha^Q(\xi)\right] > 0$, then for any such ξ*

$$\Phi_\alpha^Q(\xi) \geq \Phi_\alpha^+(\xi) \geq \Phi_\alpha(\xi) \geq \Phi_\alpha^-(\xi),$$

moreover, if there is a point $t_{\alpha,\xi'}$ such that $Pr\left[\phi(\xi',\boldsymbol{\theta}) \geq t_{\alpha,\xi'}\right] = 1-\alpha$ for some ξ', then $\Phi_\alpha^-(\xi') = \Phi_\alpha(\xi')$.

If $\Phi_\alpha^-(\xi)$ is not defined, i.e. when $Pr\left[\phi(\xi,\boldsymbol{\theta}) < \Phi_\alpha^Q(\xi)\right] = 0$, then for any such ξ

$$\Phi_\alpha^Q(\xi) = q_{\phi(\xi,\boldsymbol{\theta})}(\alpha) = \Phi_\alpha^+(\xi) = \Phi_\alpha(\xi).$$

Proof. The inequalities $\Phi_\alpha^Q(\xi) \geq \Phi_\alpha^+(\xi)$ and $\Phi_\alpha^+(\xi) \geq \Phi_\alpha^-(\xi)$ (if $\Phi_\alpha^-(\xi)$ is well-defined) are obvious from Def. 3.14, hence, first we shall prove $\Phi_\alpha^+(\xi) \geq \Phi_\alpha(\xi)$ and $\Phi_\alpha(\xi) \geq \Phi_\alpha^-(\xi)$ (if $\Phi_\alpha^-(\xi)$ is well-defined). Using the fact that $Pr\left[\phi(\xi,\boldsymbol{\theta}) \leq Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] \geq Pr\left[\phi(\xi,\boldsymbol{\theta}) \leq q_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] \geq \alpha$ and according to Consequence 3.11 and from (3.10), we have

$$\begin{aligned}
\Phi_\alpha^+(\xi) &= E\left[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) \leq Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] \\
&= \frac{E\left[\min\left\{0, \phi(\xi,\boldsymbol{\theta}) - Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right\}\right]}{Pr\left[\phi(\xi,\boldsymbol{\theta}) \leq Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]} + Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha) \\
&\geq \frac{1}{\alpha} E\left[\min\left\{0, \phi(\xi,\boldsymbol{\theta}) - Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right\}\right] + Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha) = \Phi_\alpha(\xi),
\end{aligned}$$

since the last expectation is not positive.

Similarly, using that $0 < Pr\left[\phi(\xi,\boldsymbol{\theta}) < Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] \le \alpha$ and (3.10), we obtain

$$\begin{aligned}\Phi_\alpha^-(\xi) &= E\left[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) < Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] \\ &= \frac{E\left[\min\left\{0, \phi(\xi,\boldsymbol{\theta}) - Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right\}\right]}{Pr\left[\phi(\xi,\boldsymbol{\theta}) < Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]} + Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha) \\ &\le \frac{1}{\alpha} E\left[\min\left\{0, \phi(\xi,\boldsymbol{\theta}) - Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right\}\right] + Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha) = \Phi_\alpha(\xi), \qquad (3.13)\end{aligned}$$

and the first part of the theorem is proved. Moreover, the existence of the point $t_{\alpha,\xi'}$ implies that $Pr\left[\phi(\xi',\boldsymbol{\theta}) < Rq_{\phi(\xi',\boldsymbol{\theta})}(\alpha)\right] = \alpha$ and we obtain an equality sign in (3.13) for $\xi = \xi'$.

Now let us assume that $\Phi_\alpha^-(\xi)$ is not defined, i.e. $Pr\left[\phi(\xi,\boldsymbol{\theta}) < Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] = 0$. From (3.11) it follows that

$$\begin{aligned}\Phi_\alpha(\xi) &= Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\left\{1 - \frac{Pr\left[\phi(\xi,\boldsymbol{\theta}) \le Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]}{\alpha}\right\} \\ &\quad + E\left[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) \le Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] \frac{Pr\left[\phi(\xi,\boldsymbol{\theta}) \le Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]}{\alpha} \\ &= Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\left\{1 - \frac{Pr\left[\phi(\xi,\boldsymbol{\theta}) = Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]}{\alpha}\right\} \\ &\quad + E\left[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) = Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right] \frac{Pr\left[\phi(\xi,\boldsymbol{\theta}) = Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]}{\alpha} \\ &= Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha) = \Phi_\alpha^Q(\xi).\end{aligned}$$

Hence, from previous considerations we have $\Phi_\alpha^Q(\xi) \ge \Phi_\alpha^+(\xi) \ge \Phi_\alpha(\xi) = \Phi_\alpha^Q(\xi)$ and, in addition, (3.12) implies $\Phi_\alpha^Q(\xi) \ge q_{\phi(\xi,\boldsymbol{\theta})}(\alpha) \ge \Phi_\alpha(\xi) = \Phi_\alpha^Q(\xi)$ which proves the last part of the theorem. □

We further refer the reader to Rockafellar and Uryasev (2002), who looked at jumps in the distribution function of $\phi(\xi,\boldsymbol{\theta})$, for the proofs and for comprehensive and more precise results in the risk theory.

We remind here that, generally, $\Phi_\alpha^Q(\xi)$, $\Phi_\alpha^+(\xi)$ and $\Phi_\alpha^-(\xi)$ are not concave in ξ, hence they are not easily optimized and we used them only to demonstrate the properties of the criterion $\Phi_\alpha(\xi)$. In general, the CVaR-criterion Φ_α can be considered as a compromise criterion between $\Phi_\alpha^+(\xi)$ and $\Phi_\alpha^-(\xi)$, which are nicely interpreted via the conditional expectation of $\phi(\xi,\boldsymbol{\theta})$.

Example 3.16. Consider the nonlinear regression model

$$\eta(x,\boldsymbol{\theta}) = \theta_1 \mathrm{e}^{\theta_2 x}$$

with $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ and $x \in \{-1, 1\}$. The local information matrix associated with the design $\xi = \begin{Bmatrix} -1 & 1 \\ (1-s) & s \end{Bmatrix}$, $s \in [0,1]$, is

$$M(\xi, \boldsymbol{\theta}) = s \begin{pmatrix} \mathrm{e}^{2\theta_2} & \theta_1 \mathrm{e}^{2\theta_2} \\ \theta_1 \mathrm{e}^{2\theta_2} & \theta_1^2 \mathrm{e}^{2\theta_2} \end{pmatrix} + (1-s) \begin{pmatrix} \mathrm{e}^{-2\theta_2} & -\theta_1 \mathrm{e}^{-2\theta_2} \\ -\theta_1 \mathrm{e}^{-2\theta_2} & \theta_1^2 \mathrm{e}^{-2\theta_2} \end{pmatrix}. \tag{3.14}$$

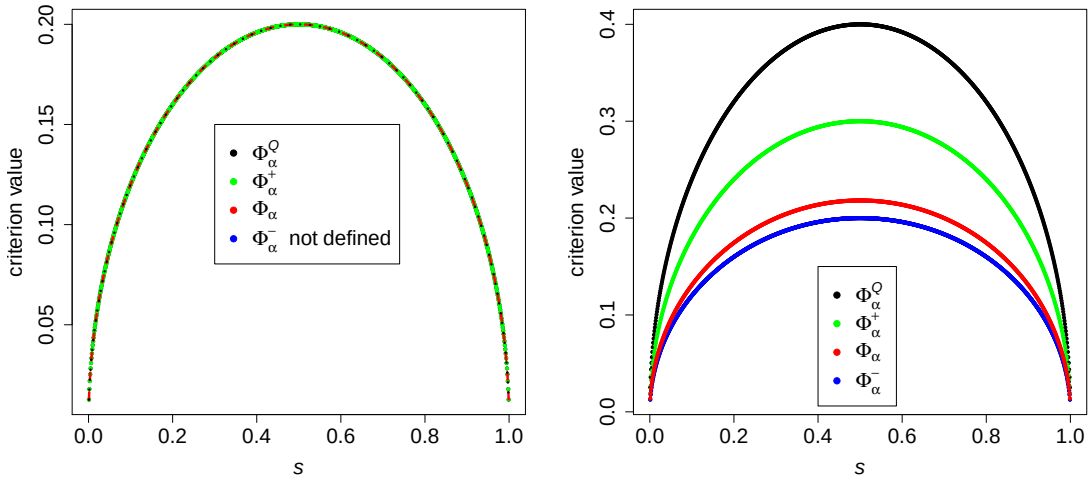
We suppose that the parameter $\boldsymbol{\theta}$ has a discrete prior distribution: θ_1 and θ_2 are independent random variables, both distributed uniformly over given finite set, θ_1 over $\{0.2, 0.4, 0.6, 0.8, 1\}$ and θ_2 over $\{-0.4, -0.2, 0.2, 0.4, 1\}$. Hence, $\pi\left(\boldsymbol{\theta}^0\right) = 1/25$ for any $\boldsymbol{\theta}^0 \in \{0.2, 0.4, 0.6, 0.8, 1\} \times \{-0.4, -0.2, 0.2, 0.4, 1\} = \Theta$. We consider the criteria of D-, A-, and E-optimality, in each case $\phi(\xi, \boldsymbol{\theta})$ is a discrete random variable.

In Figure 3.3 are displayed the graphs of $\Phi_\alpha^Q(\xi)$, $\Phi_\alpha^+(\xi)$, $\Phi_\alpha(\xi)$ and $\Phi_\alpha^-(\xi)$ as functions of s to illustrate the properties formulated in Theorem 3.15 and the missing concavity of $\Phi_\alpha^Q(\xi)$, $\Phi_\alpha^+(\xi)$, and $\Phi_\alpha^-(\xi)$.

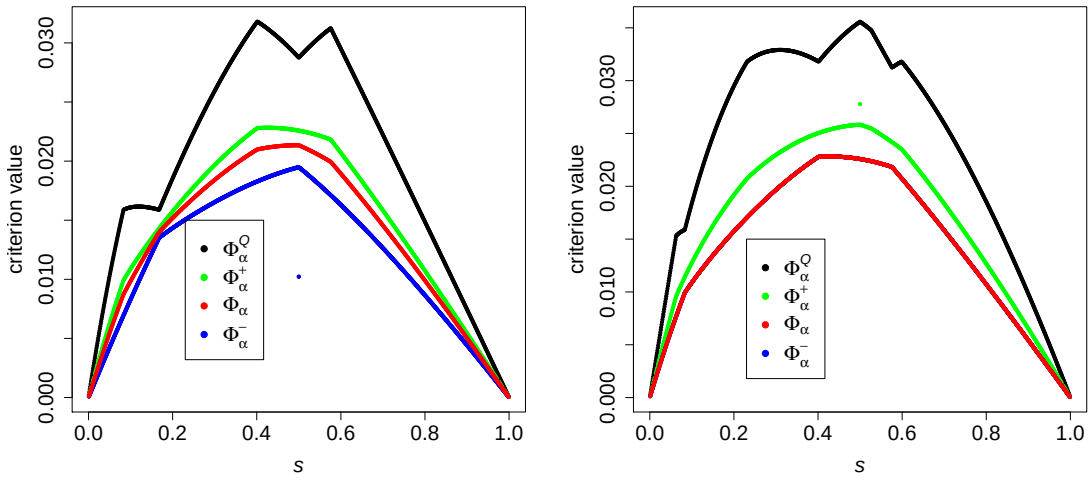
The determinant of the information matrix $M(\xi, \boldsymbol{\theta})$ in (3.14) does not depend on θ_2. Hence the D-optimality criterion attains for given ξ only five possible values, each with probability $1/5$, and for $\alpha < 1/5$ the criterion value $\Phi_\alpha^Q(\xi)$ corresponds to the minimum of these five values. This implies that for $\alpha < 1/5$ the criterion $\Phi_\alpha^-(\xi)$ is not defined and according to Theorem 3.15, $\Phi_\alpha^Q(\xi) = \Phi_\alpha^+(\xi) = \Phi_\alpha(\xi)$ as seen in Figure 3.3a. For $\alpha \geq 1/5$ the criteria follow the property $\Phi_\alpha^Q(\xi) \geq \Phi_\alpha^+(\xi) \geq \Phi_\alpha(\xi) \geq \Phi_\alpha^-(\xi)$, see Figure 3.3b.

For special choices of α such that for some $t_{\alpha,\xi} \in \mathbb{R}$ $Pr\left[\phi(\xi, \boldsymbol{\theta}) < t_{\alpha,\xi}\right] = \alpha$ (see Figures 3.3d and 3.3f), the criteria $\Phi_\alpha^-(\xi)$ and $\Phi_\alpha(\xi)$ are equal at ξ, which is also proved in Theorem 3.15.

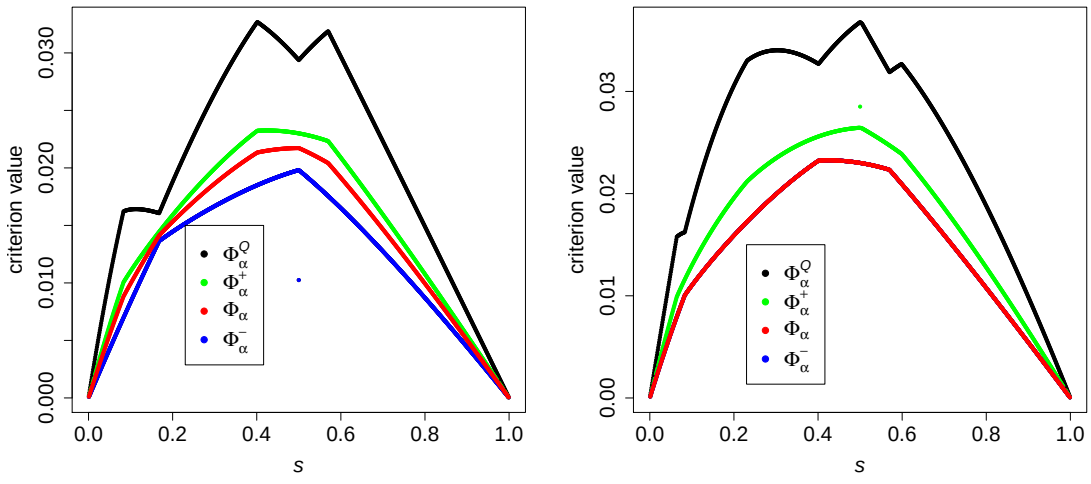
Figure 3.4 displays the $\phi_E(\xi, \boldsymbol{\theta})$-criterion values for all 25 parameters in the parametric space Θ. One sees that the criterion values for two (or more) different parameters $\boldsymbol{\theta}$ may coincide at some points s'. The graph of function $\Phi_\alpha^Q(\xi)$ contains some kinks, which appear at such points s', as seen in Figure 3.4 on the right hand side. Moreover, at these kink points the value of $\Phi_\alpha^-(\xi)$ can be significantly less compared to the neighbouring points (because less parameters satisfy the property $\phi(\xi, \boldsymbol{\theta}) < \Phi_\alpha^Q(\xi, \boldsymbol{\theta})$) or, on the other hand, $\Phi_\alpha^+(\xi)$ can be significantly greater (because more parameters satisfy the property $\phi(\xi, \boldsymbol{\theta}) \leq \Phi_\alpha^Q(\xi, \boldsymbol{\theta})$). This justifies the existence of isolated points which can be seen in Figures 3.3c–3.3f at $s = 0.5$. The isolated points appear with every kink, which, unfortunately, can not be seen in the figures due to the computational restrictions associated with displaying graphs of discontinuous functions. △



(a) D-optimality, $\alpha = 0.1$ (b) D-optimality, $\alpha = 0.22$



(c) A-optimality, $\alpha = 0.1$ (d) A-optimality, $\alpha = 0.12$



(e) E-optimality, $\alpha = 0.1$ (f) E-optimality, $\alpha = 0.12$

Figure 3.3: Example 3.16: The graphs of $\Phi_\alpha^Q(\xi)$, $\Phi_\alpha^+(\xi)$, $\Phi_\alpha(\xi)$ and $\Phi_\alpha^-(\xi)$.

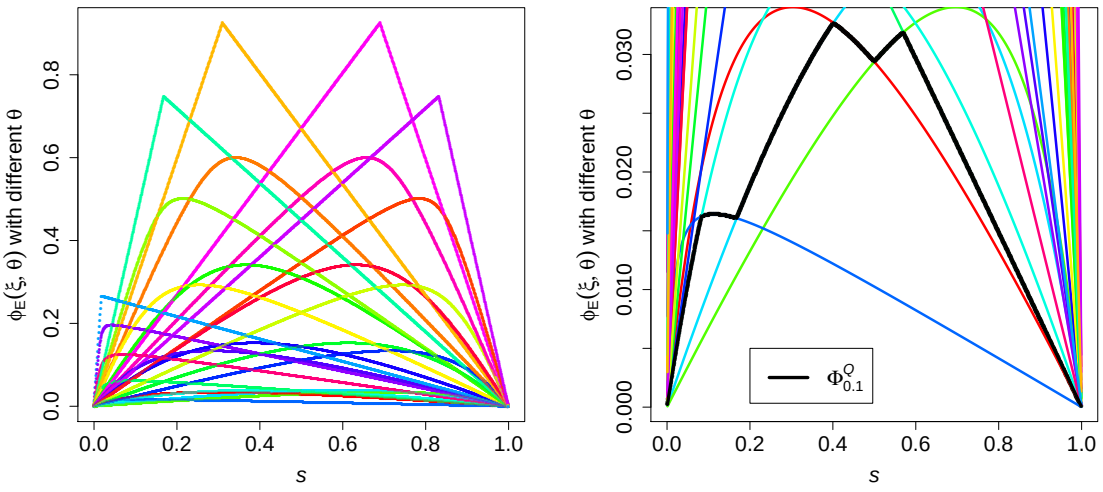


Figure 3.4: Example 3.16: The graphs of $\phi_E(\xi,\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$ in Θ (left-hand side). On the right-hand side is the enlarged part of the same figure with displayed values of the quantile criterion $\Phi^Q_{0.1}(\xi)$.

3.3.3 The CVaR-criterion in the case of continuous random variable $\phi(\xi,\boldsymbol{\theta})$

Theorem 3.17. *Assume that for $\xi\in\Xi$ is $\phi(\xi,\boldsymbol{\theta})$ a continuously distributed random variable with a cdf $F_{\phi(\xi,\boldsymbol{\theta})}$ and a pdf $f_{\phi(\xi,\boldsymbol{\theta})}$. The level $\alpha\in(0,1]$ is given. Then*

$$\Phi^+_\alpha(\xi)=\Phi_\alpha(\xi)=\Phi^-_\alpha(\xi)=\frac{1}{\alpha}\int_{-\infty}^{\Phi^Q_\alpha(\xi)} t f_{\phi(\xi,\boldsymbol{\theta})}(t)\,dt. \tag{3.15}$$

Proof. The proof follows the lines of the proof of Theorem 1 in Rockafellar and Uryasev (2000). We have only to prove $\Phi_\alpha(\xi)=\frac{1}{\alpha}\int_{-\infty}^{\Phi^Q_\alpha(\xi)} t f_{\phi(\xi,\boldsymbol{\theta})}(t)\,dt$, because from the definitions of Φ^+_α and Φ^-_α it is obvious that $\Phi^+_\alpha(\xi)=\Phi^-_\alpha(\xi)$ and Theorem 3.15 implies $\Phi^+_\alpha(\xi)=\Phi_\alpha(\xi)=\Phi^-_\alpha(\xi)$. From Consequence 3.11 and from the definition of the quantile criterion $\Phi^Q_\alpha(\xi)$ in (3.1) it follows that

$$\begin{aligned}\Phi_\alpha(\xi)&=Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)+\frac{1}{\alpha}\int_{-\infty}^{Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)}\left[t-Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]f_{\phi(\xi,\boldsymbol{\theta})}(t)\,dt\\&=Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)+\frac{1}{\alpha}\int_{-\infty}^{\Phi^Q_\alpha(\xi)} t f_{\phi(\xi,\boldsymbol{\theta})}(t)\,dt-\frac{1}{\alpha}Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\,F_{\phi(\xi,\boldsymbol{\theta})}\left[Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)\right]\\&=\frac{1}{\alpha}\int_{-\infty}^{\Phi^Q_\alpha(\xi)} t f_{\phi(\xi,\boldsymbol{\theta})}(t)\,dt.\end{aligned}$$

□

Example 3.18. Consider the nonlinear regression model $\eta(x,\theta)=e^{-x\theta}$, where $x>0$ and the one-dimensional model parameter θ has uniform prior on the closed interval

$[\theta_{\min}, \theta_{\max}]$ with $0 < \theta_{\min} < \theta_{\max}$. The elementary information matrix is $M(x,\theta) = x^2 e^{-2x\theta}$. In the example we restrict our attention to one-point designs, i.e. the experiment will consist of one single measurement in $x > 0$.

First, we considered the D-optimality criterion $\phi_D(x,\theta) = \det[M(x,\theta)] = x^2 e^{-2x\theta}$. The AVE-optimality give us the optimal x as the solution of the following optimization problem

$$x^\star_{\text{AVE}} = \arg\max_{x:x>0} \int_{\theta_{\min}}^{\theta_{\max}} \frac{\phi_D(x,\theta)}{\theta_{\max} - \theta_{\min}} d\theta = \arg\max_{x:x>0} \frac{x(e^{-2x\theta_{\min}} - e^{-2x\theta_{\max}})}{2(\theta_{\max} - \theta_{\min})},$$

and, on the other hand, the maximin criterion leads to an optimal measurement $x^\star_{\min}$, which does not depend on $\theta_{\min}$:

$$x^\star_{\min} = \frac{1}{\theta_{\max}},$$

(see Pronzato and Pázman, 2013, Examples 8.2 and 8.5). Direct calculations give that $F_{\phi_D(x,\theta)}(y) = 1 - F_\theta\left[\frac{\ln\left(\frac{y}{x^2}\right)}{-2x}\right]$ and it follows that $\Phi^Q_\alpha(x) = x^2 e^{-2x[(1-\alpha)\theta_{\max} + \alpha\theta_{\min}]}$, which is maximized at

$$x^\star_Q = \frac{1}{(1-\alpha)\theta_{\max} + \alpha\theta_{\min}}.$$

The random variable $\phi_D(x,\theta)$ has continuous distribution with density $f_{\phi_D(x,\theta)}(y) = \frac{1}{2xy(\theta_{\max} - \theta_{\min})}$ and hence

$$\begin{aligned} x^\star_{\text{CVaR}} &= \arg\max_{x:x>0} E[\phi_D(x,\theta) \mid \phi_D(x,\theta) < \Phi^Q_\alpha(x)] \\ &= \arg\max_{x:x>0} \frac{1}{\alpha} \int_{x^2 e^{-2x\theta_{\max}}}^{\Phi^Q_\alpha(x)} y f_{\phi_D(x,\theta)}(y) dy \\ &= \arg\max_{x:x>0} \frac{x\left\{e^{-2x[(1-\alpha)\theta_{\max} + \alpha\theta_{\min}]} - e^{-2x\theta_{\max}}\right\}}{2\alpha(\theta_{\max} - \theta_{\min})} \end{aligned}$$

is the value which maximizes the CVaR criterion $\Phi_\alpha(x)$.

The criterion values $\Phi_{\min}(x), \Phi_\alpha(x), \Phi^Q_\alpha(x)$, and $\Phi_{\text{AVE}}(x)$ are for $\theta_{\min} = 1/2$, $\theta_{\max} = 7/2$, $\alpha = 0.45$ and x corresponding to the quantile-optimal single-point design, i.e. $x = \frac{1}{0.55 \cdot 7/2 + 0.45 \cdot 1/2}$, depicted in Figure 3.5 on the left hand side. The displayed value $\Phi_{\min}(x)$ is the smallest possible value of the random variable $\phi_D(x,\theta)$. The area under the graph of pdf $f_{\phi_D(x,\theta)}$ between $\Phi_{\min}(x)$ and $\Phi^Q_\alpha(x)$ equals α (the gray area in the figure). In the case of continuous random variable, the CVaR criterion $\Phi_\alpha(x)$ corresponds to the expectation of values between $\Phi_{\min}(x)$ and $\Phi^Q_\alpha(x)$. Hence, by maximizing the quantile criterion we neglect the values of $\phi_D(x,\theta)$ which are smaller or larger than the quantile, while by maximizing the CVaR criterion we neglect only the values larger than the corresponding quantile.

In Figure 3.5 on the right hand side is illustrated the situation for uniform prior distribution on finite $\Theta = \{0.5, 1.25, 2, 2.75, 3.5\}$. In this case, the random variable

$\phi_D(x,\theta)$ is discrete and takes five possible values, each with probability 0.2. One sees that $Pr[\phi_D(x,\theta) < \Phi_\alpha^Q(x)] = 0.4 < \alpha$ but $Pr[\phi_D(x,\theta) \leq \Phi_\alpha^Q(x)] = 0.6 > \alpha$ and the criteria $\Phi_\alpha(x)$, $\Phi_\alpha^+(x)$ and $\Phi_\alpha^-(x)$ are no more equal.

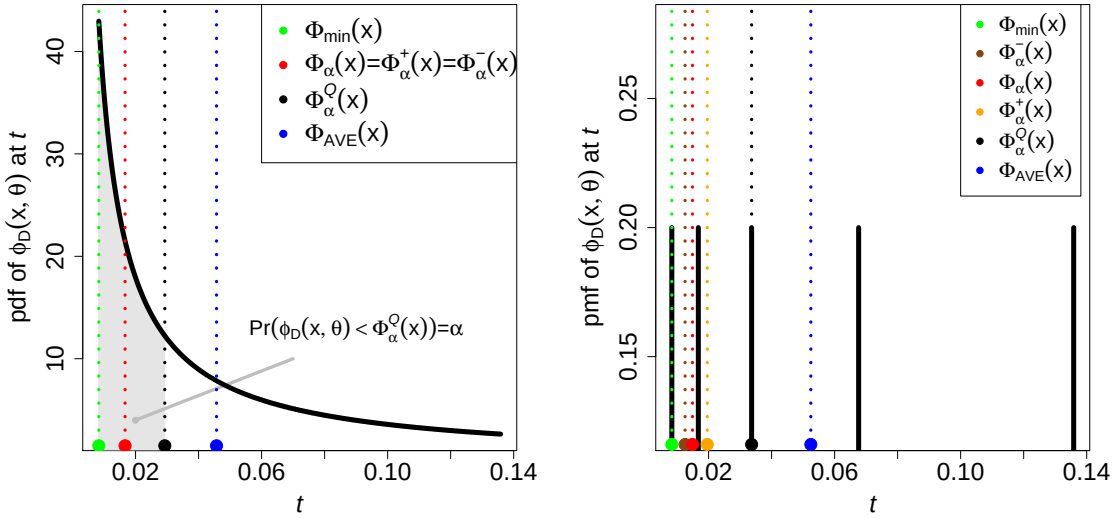


Figure 3.5: Example 3.18: The density (θ uniformly distributed on $\Theta = [1/2, 7/2]$) and pmf (θ uniformly distributed on $\Theta = \{0.5, 1.25, 2, 2.75, 3.5\}$) of $\phi_D(x,\theta)$ for $x = \frac{1}{0.55 \cdot 7/2 + 0.45 \cdot 1/2}$ with corresponding values of some criteria depicted.

Now consider the D-optimality criterion $\phi_{D*}(x,\theta) = \ln\det[M(x,\theta)]$ with θ uniformly distributed over $[\theta_{\min}, \theta_{\max}]$. The random variable $\phi_{D*}(x,\theta)$ is then uniformly distributed on the interval $[2\ln(x) - 2x\theta_{\max}, 2\ln(x) - 2x\theta_{\min}]$ and $\Phi_\alpha^Q(x) = 2\ln(x) - 2x[(1-\alpha)\theta_{\max} + \alpha\theta_{\min}]$, and hence $\Phi_\alpha(x) = E[\phi_{D*}(x,\theta) \mid \phi_{D*}(x,\theta) < \Phi_\alpha^Q(x)] = \frac{\{\ln(x) - x[(1-\alpha)\theta_{\max} + \alpha\theta_{\min}]\}^2 - [\ln(x) - x\theta_{\max}]^2}{\alpha x(\theta_{\max} - \theta_{\min})}$, which is maximized at

$$x^{\star\star}_{\text{CVaR}} = \frac{2(\theta_{\min} - \theta_{\max})}{(\alpha-2)\theta_{\max}^2 + \alpha\theta_{\min}^2 + 2(1-\alpha)\theta_{\min}\theta_{\max}}.$$

The quantile and maximin single-point optimal designs are the same as in the case of $\phi_D(x,\theta)$, since they are invariant to nonlinear rescaling of the criterion function ϕ (see Sect. 3.4.2). The AVE optimality criterion (see Pronzato and Pázman, 2013, Example 8.2) is maximized at

$$x^{\star\star}_{\text{AVE}} = \frac{1}{E(\theta)} = \frac{2}{\theta_{\min} + \theta_{\max}}.$$

One sees that only the maximin optimal design does not depend on $\theta_{\min}$ at all. △

In the case of continuously distributed $\phi(\xi, \boldsymbol{\theta})$ we obtained as a consequence of Theorem 3.17 an alternative definition of the CVaR-criterion via $\Phi_\alpha^+(\cdot)$ and $\Phi_\alpha^-(\cdot)$, and, instead of the rather abstract definition of Φ_α in (3.7), we can use an intuitively more clear definition of the CVaR-criterion via the integral in (3.15). Moreover, as a

consequence of Theorem 3.17, $\Phi_\alpha^+(\cdot)$ and $\Phi_\alpha^-(\cdot)$ are concave in ξ, which is not always true for discrete random variables.

In the simpler case when $\phi(\xi,\boldsymbol{\theta})$ is assumed to be continuous random variable, the CVaR-criterion defined in (3.7) can be formulated as

$$\Phi_\alpha(\xi) = \frac{1}{\alpha}\int_{\left\{\boldsymbol{\theta}:\phi(\xi,\boldsymbol{\theta})\leq\Phi_\alpha^Q(\xi)\right\}} \phi(\xi,\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}), \tag{3.16}$$

see also Theorem 3.17. This is the concave version of the CVaR-criterion defined in Valenzuela et al. (2015) corresponding to the definition of CVaR in Rockafellar and Uryasev (2000), where its convexity is proven for the case of continuous random variable $\phi(\xi,\boldsymbol{\theta})$ (see their Theorem 2). However, as proved in Sect. 3.2.1, the CVaR-criterion (3.7), the definition of which is based on the formula derived in Rockafellar and Uryasev (2000) and used in Pflug (2000) as definition of CVaR, is concave for any random variable. Rockafellar and Uryasev (2002) considered a more general CVaR than in Rockafellar and Uryasev (2000), as a mean of a random variable with cdf $F^\alpha_{\phi(\xi,\boldsymbol{\theta})}:\mathbb{R}\mapsto[0,1]$:

$$F^\alpha_{\phi(\xi,\boldsymbol{\theta})}(t) = \begin{cases} 1 & \text{if} \quad t \geq Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha), \\ \frac{F_{\phi(\xi,\boldsymbol{\theta})}(t)}{\alpha} & \text{otherwise,} \end{cases}$$

which is given as a proper rescaling of the cdf $F_{\phi(\xi,\boldsymbol{\theta})}$. This definition coincides with the Pflugs definition, as proved in Theorem 10 of Rockafellar and Uryasev (2002).

The reason for involving also discontinuous distributions into our considerations is that the distribution $\pi(\boldsymbol{\theta})$ can be discrete (this also includes the case when the parametric space Θ is finite or countable) causing that the distribution of $\phi(\xi,\boldsymbol{\theta})$ is discrete. However, neither the case when $\boldsymbol{\theta}$ is continuous random variable is unambiguous and simple, because the continuity of $\phi(\xi,\boldsymbol{\theta})$ is still not ensured.

3.4 Interpretation of the CVaR-criterion Φ_α in the context of experimental design

In this section we explain the main properties of the optimality criteria based on VaR and CVaR, which can be effectively applied when designing experiments in nonlinear models.

When the criterion $\phi(\xi,\boldsymbol{\theta})$ can be considered as a continuous random variable, we can formulate the CVaR-criterion $\Phi_\alpha(\xi)$ as in (3.16). So, unlike the AVE-criterion, the idea of the CVaR-criterion is to maximize the mean of $\phi(\xi,\boldsymbol{\theta})$ under the condition that $\phi(\xi,\boldsymbol{\theta}) \leq \Phi_\alpha^Q(\xi)$ (naturally, this condition is satisfied with the probability α for any ξ).

First of all, we shall compare the CVaR-criterion to the local, maximin, AVE and quantile criteria, which are more usual in design of experiments.

3.4.1 Relation to the local, maximin, and AVE-criterion

It turns out that the local, AVE and maximin criteria are special cases of the CVaR-criterion (3.7).

Theorem 3.19. *We suppose that the unknown parameter $\boldsymbol{\theta}$ has prior distribution $\pi(\boldsymbol{\theta})$ on the parametric space Θ. Then*

a) $\Phi_1(\xi) = \Phi_{\mathrm{AVE}}(\xi)$ *for any* ξ, *i.e. for* $\alpha = 1$ *the CVaR-criterion is equal to the AVE-criterion (1.10),*

b) if $\phi(\xi', \cdot) : \mathbb{R}^m \to \mathbb{R}$, $\boldsymbol{\theta} \mapsto \phi(\xi', \boldsymbol{\theta})$ *is continuous function on* $\mathbb{R}^m$ *and for every open set* $T \subset \mathbb{R}^m$ *such that* $T \cap \Theta \neq \emptyset$ *the probability* $Pr(\boldsymbol{\theta} \in T) = \pi(\Theta \cap T)$ *is positive, then, supposing that* $\Phi_{\min}(\xi')$ *does exist,* $\lim_{\alpha \to 0} \Phi_\alpha(\xi') = \Phi_{\min}(\xi')$ *, i.e. at* ξ' *the CVaR-criterion tends to the maximin criterion (1.9),*

c) if $\pi(\boldsymbol{\theta}) = 1$ *for* $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ *and* $\pi(\boldsymbol{\theta}) = 0$ *for* $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$, *then* $\Phi_\alpha(\xi) = \Phi_{\mathrm{loc}}(\xi)$ *for any* $\alpha \in (0, 1]$, *i.e. the CVaR-criterion coincides with the local criterion (1.8).*

Proof.

a) Evidently, $\min\left\{0, \phi(\xi, \boldsymbol{\theta}) - q_{\phi(\xi,\boldsymbol{\theta})}(1)\right\} = \phi(\xi, \boldsymbol{\theta}) - q_{\phi(\xi,\boldsymbol{\theta})}(1)$. So, according to Consequence 3.11, for $\alpha = 1$ the CVaR-criterion (3.7) equals $q_{\phi(\xi,\boldsymbol{\theta})}(1) + E\left[\phi(\xi, \boldsymbol{\theta}) - q_{\phi(\xi,\boldsymbol{\theta})}(1)\right] = E[\phi(\xi, \boldsymbol{\theta})] = \int_\Theta \phi(\xi, \boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$ which coincides with the AVE-criterion (1.10).

b) Let $\alpha \to 0$. We shall show that $\lim_{\alpha \to 0} \Phi_\alpha(\xi') = \min_{\boldsymbol{\theta} \in \Theta} \phi(\xi', \boldsymbol{\theta}) \equiv \Phi_{\min}(\xi')$, i.e. we have to prove that $\forall \epsilon > 0\ \exists \Delta > 0 :\ |\Phi_\delta(\xi') - \Phi_{\min}(\xi')| \leq \epsilon\ \forall 0 < \delta < \Delta$. It follows from (3.12), see also the notation in (3.8) and the definition of $\Phi_\alpha(\xi)$ in (3.9), that $\Phi_{\min}(\xi') = w_\delta[\xi', \Phi_{\min}(\xi')] \leq \max_{c \in \mathbb{R}} w_\delta(\xi', c) = \Phi_\delta(\xi') \leq \Phi_\delta^Q(\xi')$. Hence

$$0 \leq \Phi_\delta\left(\xi'\right) - \Phi_{\min}\left(\xi'\right) \leq \Phi_\delta^Q\left(\xi'\right) - \Phi_{\min}\left(\xi'\right),$$

and we have only to show that $\Phi_\delta^Q(\xi') - \Phi_{\min}(\xi') \leq \epsilon$ for any $0 < \delta < \Delta$. Take any Δ satisfying the inequality $1 - \Delta \geq Pr[\phi(\xi', \boldsymbol{\theta}) \geq \epsilon + \Phi_{\min}(\xi')]$ which is equivalent to the inequality $\Delta \leq Pr[\phi(\xi', \boldsymbol{\theta}) < \epsilon + \Phi_{\min}(\xi')]$. Indeed, for any $\epsilon > 0$ there does exist such $\Delta \in (0, 1]$, because the set $T = \{\mathbf{a} : \phi(\xi', \mathbf{a}) \in (-\infty, \epsilon + \Phi_{\min}(\xi'))\}$ is open, since it is the pre-image of the open set $(-\infty, \Phi_{\min}(\xi') + \epsilon)$ and $\phi(\xi', \cdot)$

is continuous function. Therefore, $Pr\left[\phi\left(\xi',\boldsymbol{\theta}\right) < \epsilon + \Phi_{\min}\left(\xi'\right)\right] = Pr\left(\boldsymbol{\theta} \in T\right) > 0$. From the definition of the quantile criterion (3.1), we have

$$Pr\left[\phi\left(\xi',\boldsymbol{\theta}\right) \geq \Phi_{\delta}^{Q}\left(\xi'\right)\right] \geq 1-\delta > 1-\Delta \geq Pr\left[\phi\left(\xi',\boldsymbol{\theta}\right) \geq \epsilon + \Phi_{\min}\left(\xi'\right)\right]. \quad (3.17)$$

The function $t \mapsto \Pr\left[\phi\left(\xi',\boldsymbol{\theta}\right) \geq t\right]$ is non-increasing, hence it follows from (3.17) that

$$\Phi_{\delta}^{Q}\left(\xi'\right) \leq \epsilon + \Phi_{\min}\left(\xi'\right)$$

and the statement is proved.

c) It is easy to see that $q_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right) = \phi\left(\xi,\boldsymbol{\theta}^{0}\right) = \Phi_{\text{loc}}\left(\xi\right)$ for any $\alpha \in (0,1]$. From Consequence 3.11 it follows that

$$\begin{aligned}\Phi_{\alpha}\left(\xi\right) &= q_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right) + \frac{1}{\alpha}E\left[\min\left\{0, \phi\left(\xi,\boldsymbol{\theta}\right) - q_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right)\right\}\right] = \\ &= q_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right) + \frac{1}{\alpha}\left[\min\left\{0, \phi\left(\xi,\boldsymbol{\theta}^{0}\right) - q_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right)\right\}\right] = q_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right) = \Phi_{\text{loc}}\left(\xi\right).\end{aligned}$$

□

3.4.2 Relation to the quantile criterion

The quantile criterion keeps different ordering of designs than the CVaR-criterion.

Pázman and Pronzato (2007) emphasize that the quantile criterion is invariant to nonlinear rescaling of the criterion $\phi(\xi,\boldsymbol{\theta})$. Indeed, unlike the AVE- and the CVaR-criterion, when taking an increasing function $\varrho : \mathbb{R} \to \mathbb{R}$, then

$$\begin{aligned}Rq_{\varrho[\phi(\xi,\boldsymbol{\theta})]}\left(\alpha\right) &= \sup\left\{t \in \mathbb{R} : Pr\left\{\varrho\left[\phi\left(\xi,\boldsymbol{\theta}\right)\right] \geq t\right\} \geq 1-\alpha\right\} \\ &= \sup\left\{\varrho\left[\varrho^{-1}\left(t\right)\right] \in \mathbb{R} : Pr\left[\phi\left(\xi,\boldsymbol{\theta}\right) \geq \varrho^{-1}\left(t\right)\right] \geq 1-\alpha\right\} \\ &= \sup\left\{\varrho\left(t\right) \in \mathbb{R} : Pr\left[\phi\left(\xi,\boldsymbol{\theta}\right) \geq t\right] \geq 1-\alpha\right\} \\ &= \varrho\left[Rq_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right)\right] = \varrho\left[\Phi_{\alpha}^{Q}\left(\xi\right)\right],\end{aligned} \quad (3.18)$$

hence the ordering of designs given by $\Phi_{\alpha}^{Q}\left(\xi\right)$ is the same regardless of the transformation of the initial criterion $\phi\left(\xi,\boldsymbol{\theta}\right)$.

On the other hand,

$$\begin{aligned}&E\left\{\varrho\left[\phi\left(\xi,\boldsymbol{\theta}\right)\right] \mid \varrho\left[\phi\left(\xi,\boldsymbol{\theta}\right)\right] < Rq_{\varrho[\phi(\xi,\boldsymbol{\theta})]}\left(\alpha\right)\right\} \\ =&E\left\{\varrho\left[\phi\left(\xi,\boldsymbol{\theta}\right)\right] \mid \varrho\left[\phi\left(\xi,\boldsymbol{\theta}\right)\right] < \varrho\left[Rq_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right)\right]\right\} \\ =&E\left\{\varrho\left[\phi\left(\xi,\boldsymbol{\theta}\right)\right] \mid \phi\left(\xi,\boldsymbol{\theta}\right) < Rq_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right)\right\},\end{aligned}$$

which is not necessarily equal to $\varrho\left\{E\left[\phi\left(\xi,\boldsymbol{\theta}\right) \mid \phi\left(\xi,\boldsymbol{\theta}\right) < Rq_{\phi(\xi,\boldsymbol{\theta})}\left(\alpha\right)\right]\right\} = \varrho\left[\Phi_{\alpha}\left(\xi\right)\right]$ (the equality appears when $\varrho(\cdot)$ is linear) and the CVaR-criterion for $\varrho\left[\phi\left(\xi,\boldsymbol{\theta}\right)\right]$ is not generally equal to the CVaR-criterion for $\phi\left(\xi,\boldsymbol{\theta}\right)$.

We justified in Sect. 3.2.1 that the criterion $\Phi_\alpha(\xi)$ is concave in ξ when the criterion $\phi(\xi,\boldsymbol{\theta})$ is concave in ξ for any $\boldsymbol{\theta}\in\Theta$. This is a big advantage comparing to the quantile criterion, because it can be optimized applying well-known methods of mathematical programming (see Rockafellar and Uryasev, 2000, Valenzuela et al., 2015, or Algorithm 3.22 in Sect. 3.5).

3.4.3 Further properties of the CVaR-criterion

Let the criterion $\phi(\xi,\boldsymbol{\theta})$ be positively homogeneous for any $\boldsymbol{\theta}$, then

$$\begin{aligned}\Phi_\alpha(a\xi) &= \max_{c\in\mathbb{R}}\left\{c+\frac{1}{\alpha}E\left[\min\left\{a\phi(\xi,\boldsymbol{\theta})-c,0\right\}\right]\right\}\\ &= aRq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)+\frac{1}{\alpha}E\left[\min\left\{a\phi(\xi,\boldsymbol{\theta})-aRq_{\phi(\xi,\boldsymbol{\theta})}(\alpha),0\right\}\right]\\ &= a\left\{Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)+\frac{1}{\alpha}E\left[\min\left\{\phi(\xi,\boldsymbol{\theta})-Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha),0\right\}\right]\right\}\\ &= a\max_{c\in\mathbb{R}}\left\{c+\frac{1}{\alpha}E\left[\min\left\{\phi(\xi,\boldsymbol{\theta})-c,0\right\}\right]\right\} = a\Phi_\alpha(\xi),\end{aligned}$$

where we used Consequence 3.11 and the property $aRq_Y(\alpha)=Rq_{aY}(\alpha)$ for $\alpha>0$, similarly as in (3.18), to prove the positive homogeneity of $\Phi_\alpha(\xi)$.

The directional derivative and the equivalence theorem

We could see in Figure 3.3 of Example 3.16 that the CVaR criterion is not necessarily differentiable function of the design ξ even if the criterion $\phi(\xi,\boldsymbol{\theta})$ is (e.g. A-optimality). However, the CVaR criterion is concave and hence the directional derivative at ξ in the direction ν always exists and is defined as follows

$$\mathcal{F}_{\Phi_\alpha}(\xi,\nu)=\lim_{a\to 0^+}\frac{\Phi_\alpha[(1-a)\xi+a\nu]-\Phi_\alpha(\xi)}{a},$$

see e.g. Lemma 5.16 in Pronzato and Pázman (2013). The equivalence theorem states that the design $\xi^\star$ is CVaR optimal if and only if $\sup_{\nu\in\Xi}\mathcal{F}_{\Phi_\alpha}(\xi^\star,\nu)=0$, see e.g. Theorem 5.21 in Pronzato and Pázman (2013).

Lemma 3.20. *Suppose that $\phi(\zeta,\boldsymbol{\theta})$ is for any $\boldsymbol{\theta}\in\Theta$ concave and continuous in ζ. Let $\phi(\xi,\boldsymbol{\theta})>-\infty$ for any $\boldsymbol{\theta}\in\Theta$, $E\left[|\phi(\xi,\boldsymbol{\theta})|\right]<\infty$, and let $c,b\in\mathbb{R}$. The directional derivative of the function $w_\alpha(\xi,c)=c+\frac{1}{\alpha}E\left[\min\{0,\phi(\xi,\boldsymbol{\theta})-c\}\right]$ at (ξ,c) in the direction (ν,b) equals*

$$\mathcal{F}_{w_\alpha}\left[(\xi,c),(\nu,b)\right]=b-c+\frac{1}{\alpha}E\left[\begin{cases}0 & \text{if } \phi(\xi,\boldsymbol{\theta})>c\\ \min\{0,\mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu)-(b-c)\} & \text{if } \phi(\xi,\boldsymbol{\theta})=c\\ \mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu)-(b-c) & \text{if } \phi(\xi,\boldsymbol{\theta})<c\end{cases}\right],$$

where $\mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu) = \lim_{a\to 0^+} \frac{\phi[(1-a)\xi+a\nu,\boldsymbol{\theta}]-\phi(\xi,\boldsymbol{\theta})}{a}$ *denotes the directional derivative of* $\phi(\cdot,\boldsymbol{\theta})$ *at* ξ *and in the direction* ν *and* $\alpha \in (0,1]$.

Proof. The directional derivative of the function $w_\alpha(\xi,c)$ does exist due to its concavity in (ξ,c), see Lemma 3.5b. For a fixed $\boldsymbol{\theta}$ let us use the following notation

$$h_{\boldsymbol{\theta}}(\xi,c) \equiv \phi(\xi,\boldsymbol{\theta}) - c,$$
$$g_{\boldsymbol{\theta}}(\xi,c) \equiv \min\{0, h_{\boldsymbol{\theta}}(\xi,c)\},$$

and for any $a \in [0,1]$ define

$$\xi_a = (1-a)\xi + a\nu,$$
$$c_a = (1-a)c + ab,$$

where $\xi,\nu \in \Xi$ and $c,b \in \mathbb{R}$ are fixed. Now we will evaluate the directional derivative of the function $g_{\boldsymbol{\theta}}(\cdot,\cdot)$ at (ξ,c) in the direction (ν,b), i.e.

$$\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)] = \lim_{a\to 0^+} \frac{g_{\boldsymbol{\theta}}(\xi_a,c_a) - g_{\boldsymbol{\theta}}(\xi,c)}{a}.$$

The values of ξ,ν,c,b always result in one of the following possibilities and directly determine $\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)]$:

a) *The case when* $h_{\boldsymbol{\theta}}(\xi,c) > 0$. Then thanks to the continuity of function $h_{\boldsymbol{\theta}}(\cdot,\cdot)$ there is a number $a^\star \in (0,1]$ such that $h_{\boldsymbol{\theta}}(\xi_a,c_a) > 0$ for any $a \in [0,a^\star)$, which implies that $g_{\boldsymbol{\theta}}(\xi,c) = g_{\boldsymbol{\theta}}(\xi_a,c_a) = 0$ and

$$\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)] = 0.$$

b) *The case when* $h_{\boldsymbol{\theta}}(\xi,c) < 0$. Applying the continuity of $h_{\boldsymbol{\theta}}(\cdot,\cdot)$ as in the case a) one obtains that there is a number $a^\star \in (0,1]$ such that $h_{\boldsymbol{\theta}}(\xi_a,c_a) < 0$ for any $a \in [0,a^\star)$, which implies that $g_{\boldsymbol{\theta}}(\xi,c) = h_{\boldsymbol{\theta}}(\xi,c)$ and $g_{\boldsymbol{\theta}}(\xi_a,c_a) = h_{\boldsymbol{\theta}}(\xi_a,c_a)$. It follows

$$\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)] = \lim_{a\to 0^+} \frac{\phi(\xi_a,\boldsymbol{\theta}) - c_a - \phi(\xi,\boldsymbol{\theta}) + c}{a} = \mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu) - (b-c).$$

c) *The case when* $h_{\boldsymbol{\theta}}(\xi,c) = 0$ *and* $h_{\boldsymbol{\theta}}(\xi_a,c_a) < 0$ *for any* $a \in (0,1]$. Then $g_{\boldsymbol{\theta}}(\xi,c) = h_{\boldsymbol{\theta}}(\xi,c)$ and $g_{\boldsymbol{\theta}}(\xi_a,c_a) = h_{\boldsymbol{\theta}}(\xi_a,c_a)$ like in the case b). It follows

$$\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)] = \lim_{a\to 0^+} \frac{h_{\boldsymbol{\theta}}(\xi_a,c_a) - h_{\boldsymbol{\theta}}(\xi,c)}{a} = \mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu) - (b-c) \le 0.$$

d) *The case when* $h_{\boldsymbol{\theta}}(\xi,c) = 0$ *and* $h_{\boldsymbol{\theta}}(\xi_{a^\star},c_{a^\star}) \ge 0$ *for some* $a^\star \in (0,1]$. Then due to the concavity of the function $h_{\boldsymbol{\theta}}(\cdot,\cdot)$ we have that $h_{\boldsymbol{\theta}}(\xi_a,c_a) \ge 0$ for any $a \in [0,a^\star]$. It follows, like in the case a), that

$$\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)] = 0,$$

and moreover

$$\mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu)-(b-c)=\lim_{a\to 0^+}\frac{h_{\boldsymbol{\theta}}(\xi_a,c_a)-h_{\boldsymbol{\theta}}(\xi,c)}{a}=\lim_{a\to 0^+}\frac{h_{\boldsymbol{\theta}}(\xi_a,c_a)}{a}\geq 0.$$

Summarizing, we have for a fixed $\boldsymbol{\theta}$

$$\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)]=\begin{cases}0 & \text{if } \phi(\xi,\boldsymbol{\theta})>c,\\ \min\{0,\mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu)-(b-c)\} & \text{if } \phi(\xi,\boldsymbol{\theta})=c,\\ \mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi,\nu)-(b-c) & \text{if } \phi(\xi,\boldsymbol{\theta})<c.\end{cases}$$

Assuming that the integration and limit are interchangeable (see Pronzato and Pázman, 2013, p. 237) one obtains the directional derivative of w_α at (ξ,c) in the direction (ν,b):

$$\lim_{a\to 0^+}\frac{(1-a)c+ab-c}{a}+\frac{1}{\alpha}E\left\{\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)]\right\}=b-c+\frac{1}{\alpha}E\left\{\mathcal{F}_{g_{\boldsymbol{\theta}}}[(\xi,c),(\nu,b)]\right\}.$$

□

Since $\Phi_\alpha(\xi)=w_\alpha\left[\xi,\Phi_\alpha^Q(\xi)\right]$ and $w_\alpha(\xi,c)$ is concave function in (ξ,c), see Lemma 3.5b, we can use the same reasoning as in Theorem 5.21 of Pronzato and Pázman (2013), and, the design $\xi^\star$ is CVaR optimal if and only if

$$\sup_{\substack{\nu\in\Xi\\ b\in\mathbb{R}}}\mathcal{F}_{w_\alpha}\left[\left(\xi^\star,\Phi_\alpha^Q(\xi^\star)\right),(\nu,b)\right]=0. \tag{3.19}$$

However, when the criterion $\phi(\xi,\boldsymbol{\theta})$ is differentiable for any $\boldsymbol{\theta}$ and when $\phi(\xi^\star,\boldsymbol{\theta})$ is continuous random variable, we can apply the following equivalence theorem.

Theorem 3.21. *Consider the criterion $M\mapsto\bar{\phi}(M)$ and $\bar{\phi}[M(\xi,\boldsymbol{\theta})]=\phi(\xi,\boldsymbol{\theta})$ $\forall\xi\in\Xi$, $\forall\boldsymbol{\theta}\in\Theta$. Suppose that $\bar{\phi}$ is differentiable and denote by $G(\xi,\boldsymbol{\theta})\in\mathbb{R}^{m\times m}$ its gradient with respect to $M\in\mathbb{R}^{m\times m}$ at $M(\xi,\boldsymbol{\theta})$ and let $k(\xi,\boldsymbol{\theta})=\operatorname{tr}[M(\xi,\boldsymbol{\theta})G(\xi,\boldsymbol{\theta})]$. Let $\phi(\xi^\star,\boldsymbol{\theta})$ be a continuous random variable with $E[|\phi(\xi^\star,\boldsymbol{\theta})|]<\infty$. The design $\xi^\star$ is then CVaR-optimal if and only if*

$$0=\max_{\substack{\mathbf{x}\in\mathcal{X}\\ b\in\mathbb{R}}}\left[b-\Phi_\alpha^Q(\xi^\star)+\right.$$
$$\left.+\frac{1}{\alpha}E\left(\begin{cases}0 & \text{if } \phi(\xi^\star,\boldsymbol{\theta})>\Phi_\alpha^Q(\xi^\star)\\ \operatorname{tr}[M(\mathbf{x},\boldsymbol{\theta})G(\xi^\star,\boldsymbol{\theta})]-k(\xi^\star,\boldsymbol{\theta})-(b-\Phi_\alpha^Q(\xi^\star)) & \text{if } \phi(\xi^\star,\boldsymbol{\theta})<\Phi_\alpha^Q(\xi^\star)\end{cases}\right)\right].$$

Proof. Since $\phi(\xi,\boldsymbol{\theta})$ is differentiable, one has

$$\begin{aligned}\mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi^\star,\nu)&=\operatorname{tr}[M(\nu,\boldsymbol{\theta})G(\xi^\star,\boldsymbol{\theta})]-k(\xi^\star,\boldsymbol{\theta})\\ &=\int_{\mathbf{x}\in\mathcal{X}}\operatorname{tr}[M(\mathbf{x},\boldsymbol{\theta})G(\xi^\star,\boldsymbol{\theta})]\,d\nu(\mathbf{x})-k(\xi^\star,\boldsymbol{\theta}),\end{aligned}$$

see e.g. Eq. 5.34 in Pronzato and Pázman (2013). As a consequence of differentiability, $\phi(\cdot,\boldsymbol{\theta})$ is necessarily continuous at $\xi^\star$ for any $\boldsymbol{\theta}$ and hence we may use Lemma 3.20. In addition, $Pr[\phi(\xi^\star,\boldsymbol{\theta})=c]=0$, since $\phi(\xi^\star,\boldsymbol{\theta})$ is continuous random variable.

One has for any $\nu\in\Xi$ and $c,b\in\mathbb{R}$

$$
\begin{aligned}
&\mathcal{F}_{w_\alpha}[(\xi^\star,c),(\nu,b)]-(b-c)\\
&=\frac{1}{\alpha}E\left[\begin{cases}0 & \text{if } \phi(\xi^\star,\boldsymbol{\theta})>c\\ \mathcal{F}_{\phi(\cdot,\boldsymbol{\theta})}(\xi^\star,\nu)-(b-c) & \text{if } \phi(\xi^\star,\boldsymbol{\theta})<c\end{cases}\right]\\
&=\frac{1}{\alpha}\int_{\mathbf{x}\in\mathcal{X}}E\left[\begin{cases}0 & \text{if } \phi(\xi^\star,\boldsymbol{\theta})>c\\ \operatorname{tr}[M(\mathbf{x},\boldsymbol{\theta})G(\xi^\star,\boldsymbol{\theta})]-k(\xi^\star,\boldsymbol{\theta})-(b-c) & \text{if } \phi(\xi^\star,\boldsymbol{\theta})<c\end{cases}\right]d\nu(\mathbf{x})\\
&\leq\frac{1}{\alpha}\max_{\mathbf{x}\in\mathcal{X}}E\left[\begin{cases}0 & \text{if } \phi(\xi^\star,\boldsymbol{\theta})>c\\ \operatorname{tr}[M(\delta_{\mathbf{x}},\boldsymbol{\theta})G(\xi^\star,\boldsymbol{\theta})]-k(\xi^\star,\boldsymbol{\theta})-(b-c) & \text{if } \phi(\xi^\star,\boldsymbol{\theta})<c\end{cases}\right]\\
&\leq\sup_{\zeta\in\Xi}\frac{1}{\alpha}E\left[\begin{cases}0 & \text{if } \phi(\xi^\star,\boldsymbol{\theta})>c\\ \operatorname{tr}[M(\zeta,\boldsymbol{\theta})G(\xi^\star,\boldsymbol{\theta})]-k(\xi^\star,\boldsymbol{\theta})-(b-c) & \text{if } \phi(\xi^\star,\boldsymbol{\theta})<c\end{cases}\right]\\
&=\sup_{\zeta\in\Xi}\mathcal{F}_{w_\alpha}[(\xi^\star,c),(\zeta,b)]-(b-c),
\end{aligned}
$$

where $\delta_{\mathbf{x}}$ denotes Dirac measure concentrated at $\mathbf{x}$. Since the previous inequalities are satisfied for any $\nu\in\Xi$, then they are also satisfied for the supremum over ν, and we obtain

$$
\begin{aligned}
&\sup_{\substack{\nu\in\Xi\\ b\in\mathbb{R}}}\mathcal{F}_{w_\alpha}[(\xi^\star,c),(\nu,b)]=\\
&=\max_{\substack{\mathbf{x}\in\mathcal{X}\\ b\in\mathbb{R}}}\left[b-c+\frac{1}{\alpha}E\left(\begin{cases}0 & \text{if } \phi(\xi^\star,\boldsymbol{\theta})>c\\ \operatorname{tr}[M(\mathbf{x},\boldsymbol{\theta})G(\xi^\star,\boldsymbol{\theta})]-k(\xi^\star,\boldsymbol{\theta})-(b-c) & \text{if } \phi(\xi^\star,\boldsymbol{\theta})<c\end{cases}\right)\right],
\end{aligned}
$$

which, as follows from (3.19), proves the statement for $c=\Phi_\alpha^Q(\xi^\star)$. □

3.5 Calculation of CVaR-optimal designs

The design $\xi^\star$ is CVaR-optimal if

$$
\Phi_\alpha(\xi^\star)=\max_{\xi\in\Xi}\Phi_\alpha(\xi)=\max_{\xi,c}w_\alpha(\xi,c), \tag{3.20}
$$

for the second equality see Theorem 2 in Rockafellar and Uryasev (2000) or Theorem 14 in Rockafellar and Uryasev (2002). We suggest to use the cutting plane method (see Sect. 1.4) to solve the problem (3.20).

3.5.1 Subgradient for the function $w_\alpha(\xi, c)$ from (3.9)

For the purposes of the optimization of the CVaR-criterion Φ_α via the cutting plane method, either the design space $\mathcal{X}$ has to be finite or we need to use a discretization $\mathcal{X}'$ of $\mathcal{X}$ (see examples in Sect. 3.6). Consequently the (discretized) design space consists of $\mathrm{card}(\mathcal{X}')$ points and hence the design ξ corresponds to the relative frequencies of measurements taken in these points, i.e. $\xi = (\xi(\mathbf{x}); \mathbf{x} \in \mathcal{X}')$, and can be interpreted as $\mathrm{card}(\mathcal{X}')$-dimensional real vector with nonnegative components summing to one. Moreover, we assume that the criterion $\phi(\xi, \boldsymbol{\theta})$ is concave in ξ with finite expectation $E[|\phi(\xi, \boldsymbol{\theta})|]$ for any ξ.

Now denote $\mathbf{z} = \left(\xi^\top, c\right)^\top$, $\xi \in \Xi \subset \mathbb{R}^{\mathrm{card}(\mathcal{X}')}$, $c \in \mathbb{R}$, and

$$\varphi(\mathbf{z}) = w_\alpha(\xi, c) = c + \frac{1}{\alpha} E[\min\{0, \phi(\xi, \boldsymbol{\theta}) - c\}] \tag{3.21}$$

is the function to be maximized over c and ξ. The function $w_\alpha(\xi, c)$ is concave, see Lemma 3.5b, and hence the method of cutting planes can be applied to solve the optimization problem

$$(\xi^\star, c^\star) = \arg\max_{\xi, c} w_\alpha(\xi, c) = \max_{\mathbf{z}} \varphi(\mathbf{z}). \tag{3.22}$$

To be able to use the cutting plane method, we have to evaluate the formula for subgradient. Denote by $\nabla_\xi \phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right)$ an arbitrary subgradient of the criterion $\phi(\xi, \boldsymbol{\theta})$ at $\widetilde{\xi}$. We will assume that it is known (when the criterion is differentiable, it equals to the gradient of the criterion). Using the properties from Lemma 1.4 we evaluate a

subgradient of $\varphi(\cdot)$ at $\widetilde{\mathbf{z}} = \left(\widetilde{\xi}, \widetilde{c}\right)^\top$:

$$\begin{aligned}
\nabla\varphi(\widetilde{\mathbf{z}}) = \nabla w_\alpha\left(\widetilde{\xi}, \widetilde{c}\right) =& \nabla\widetilde{c} + \frac{1}{\alpha} E\left[\nabla \min\left\{\phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) - \widetilde{c}, 0\right\}\right] \\
&= \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} + \frac{1}{\alpha} E\left[\begin{cases} \nabla 0 & \text{if } \phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) > \widetilde{c}, \\ \nabla\phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) + \nabla\left(-\widetilde{c}\right) & \text{otherwise} \end{cases}\right] \\
&= \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} + \frac{1}{\alpha} E\left[\begin{cases} \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix} & \text{if } \phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) > \widetilde{c}, \\ \begin{pmatrix} \nabla_\xi\phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ -1 \end{pmatrix} & \text{otherwise} \end{cases}\right] \\
&= \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} + \frac{1}{\alpha} E\left[\begin{cases} \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix} & \text{if } \phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) > \widetilde{c}, \\ \begin{pmatrix} \nabla_\xi\phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) \\ -1 \end{pmatrix} & \text{otherwise} \end{cases}\right].
\end{aligned} \tag{3.23}$$

Kelley (1960) proposed the cutting plane method for maximizing a continuous concave function defined on a compact convex set. In our case, the continuity of $\varphi(\mathbf{z}) = w_\alpha(\xi, c)$ is ensured on every closed bounded subset of $\mathbb{R}^{\mathrm{card}(\mathcal{X}')+1}$ in Lemma 3.9. To satisfy the requirements on compactness, we have to consider some bounds on c, say c_{low} and c_{up}. When $\phi(\xi, \cdot)$ is an isotonic criterion, the upper bound may be (pessimistically) chosen as $c_{up} = \max_{\boldsymbol{\theta}\in\Theta} \phi(\mathbf{1}, \boldsymbol{\theta})$, since for every $\boldsymbol{\theta} \in \Theta$ $\phi(\xi^\star, \boldsymbol{\theta}) \leq \phi(\mathbf{1}, \boldsymbol{\theta})$ and $c^\star$ is less than or equal to the α-right-quantile, hence it can be bounded from above by maximal possible value $\max_{\boldsymbol{\theta}\in\Theta} \phi(\mathbf{1}, \boldsymbol{\theta})$. Supposing that $\xi^{(0)} \in \mathbb{R}^{\mathrm{card}(\mathcal{X}')}$ and $c^{(0)} \in \mathbb{R}$ are starting points, then one may choose c_{low} as $w_\alpha\left(\xi^{(0)}, c^{(0)}\right)$ since $w_\alpha\left(\xi^{(0)}, c^{(0)}\right) \leq w_\alpha(\xi^\star, c^\star) \leq c^\star$ (the last inequality follows from the definition of $w_\alpha(\xi, c)$ in Eq. (3.8)), where $\mathbf{z}^\star = (\xi^\star, c^\star)^\top$ is an optimal solution of the problem (3.22). So, finally, one can approach the solution of the problem (3.20) via the method of cutting planes (1.12), i.e. by solving a sequence of linear programming problems:

$$\max\left(\mathbf{0}^\top, 0, 1\right)\begin{pmatrix}\xi^{(i+1)}\\ c^{(i+1)}\\ t^{(i+1)}\end{pmatrix} \tag{3.24}$$

$$\text{s.t.} \qquad \xi^{(i+1)} \geq \mathbf{0}, c^{(i+1)} \in [c_{low}, c_{up}] \tag{3.25}$$

$$\left(\mathbf{1}^\top, 0, 0\right)\begin{pmatrix}\xi^{(i+1)}\\ c^{(i+1)}\\ t^{(i+1)}\end{pmatrix} = 1 \tag{3.26}$$

$$\begin{pmatrix}\nabla^\top\varphi\left(\mathbf{z}^{(0)}\right) & -1\\ \vdots & \vdots\\ \nabla^\top\varphi\left(\mathbf{z}^{(i)}\right) & -1\end{pmatrix}\begin{pmatrix}\mathbf{z}^{(i+1)}\\ t^{(i+1)}\end{pmatrix} \geq \begin{pmatrix}\nabla^\top\varphi\left(\mathbf{z}^{(0)}\right)\mathbf{z}^{(0)} - \varphi\left(\mathbf{z}^{(0)}\right)\\ \vdots\\ \nabla^\top\varphi\left(\mathbf{z}^{(i)}\right)\mathbf{z}^{(i)} - \varphi\left(\mathbf{z}^{(i)}\right)\end{pmatrix} \tag{3.27}$$

(with $\mathbf{z}^{(h)} = \left(\xi^{(h)^\top}, c^{(h)}\right)^\top$, $h = 1, \ldots, i+1$) until $t^{(i+1)} - \max_{j\in\{0,\ldots,i+1\}}\varphi\left(\mathbf{z}^{(j)}\right) < \epsilon$.

Note that omitting the constraints in (3.25) may cause that the LP problem (3.24)–(3.27) is not bounded.

3.5.2 The algorithm

After evaluating the required subgradients, we are able to formulate the algorithm which leads to ϵ-optimal designs $\xi_\epsilon^\star$ maximizing the criterion $\Phi_\alpha(\xi)$. As a by-product of the algorithm we obtain the value $c_\epsilon^\star$ which maximizes the function $w_{\alpha,\phi(\xi_\epsilon^\star,\boldsymbol{\theta})}(c) = c + \frac{1}{\alpha}E\left[\min\left\{0, \phi(\xi_\epsilon^\star, \boldsymbol{\theta}) - c\right\}\right]$ and, according to Theorem 3.10, $c_\epsilon^\star$ is from the interval $\left[q_{\phi(\xi_\epsilon^\star,\boldsymbol{\theta})}, Rq_{\phi(\xi_\epsilon^\star,\boldsymbol{\theta})}\right]$.

Algorithm 3.22.

0. i Denote $\mathcal{X}'$ the discretization of the design space $\mathcal{X}$.

   ii Choose the starting design $\xi^{(0)} \in \mathbb{R}^{\text{card}(\mathcal{X}')}$, s.t. $\xi^{(0)}(\mathbf{x}) \geq 0\ \forall \mathbf{x} \in \mathcal{X}'$ and $\mathbf{1}^\top\xi^{(0)} = 1$.

   iii Choose $c_{low} \leq c_{up}$, the bounds on c.

   iv Set $\mathbf{z}^{(0)} = \left(\xi^{(0)^\top}, c^{(0)}\right)^\top$, where $c^{(0)} \in [c_{low}, c_{up}]$.

   v Take ϵ greater than 0 but small.

   vi Set $i = 0$.

1. Solve the LP problem (3.24–3.27).

2. i Set $\mathbf{z}' = \left(\xi'^\top, c'\right)^\top = \arg\max_{\mathbf{z}\in\left\{\mathbf{z}^{(0)},\mathbf{z}^{(1)},\ldots,\mathbf{z}^{(i+1)}\right\}}\varphi(\mathbf{z})$.

ii If $t^{(i+1)} - \varphi(\mathbf{z}') < \epsilon$, stop and return $\xi_\epsilon^\star = \xi'$ as ϵ-optimal design and $c_\epsilon^\star = c'$ as ϵ-optimal value which maximizes the function $w_{\alpha,\phi(\xi_\epsilon^\star,\boldsymbol{\theta})}(c)$.

iii Else set $i \leftarrow i+1$ and continue from the Step 1.

The discretization of the design space in Step 0.i of Algorithm 3.22 may influence the results. Hence, we may run Algorithm 3.22 several times, and every time enrich the discretized design space $\mathcal{X}'$ with some additional design points adjacent to the support points obtained from previous runs of Algorithm 3.22, similarly as in Example 3 of Pázman and Pronzato (2014), see also Example 3.24. Notice that as the cardinality of $\mathcal{X}'$ increases the computational time is larger.

The stopping rule (Step 2.ii) follows from the subgradient inequality (1.13). Denote $\mathbf{z}^\star$ the true solution of the optimization problem (1.11), and, according to (1.13),

$$\begin{aligned}\max_{j\in\{0,1,\ldots,i+1\}} \varphi\left(\mathbf{z}^{(j)}\right) &\leq \varphi(\mathbf{z}^\star) \\ &\leq \min_{j\in\{0,\ldots,i\}} \varphi\left(\mathbf{z}^{(j)}\right) + \nabla^\top \varphi\left(\mathbf{z}^{(j)}\right)\left(\mathbf{z}^\star - \mathbf{z}^{(j)}\right) \\ &\leq \max_{\mathbf{z}} \min_{j\in\{0,\ldots,i\}} \varphi\left(\mathbf{z}^{(j)}\right) + \nabla^\top \varphi\left(\mathbf{z}^{(j)}\right)\left(\mathbf{z} - \mathbf{z}^{(j)}\right) = t^{(i+1)}.\end{aligned}$$

When the expectations in $\varphi(\mathbf{z})$ and $\nabla\varphi(\mathbf{z})$ are computed exactly, then the stopping rule ensures that

$$\Phi_\alpha(\xi^\star) - \Phi_\alpha(\xi_\epsilon^\star) < \epsilon, \tag{3.28}$$

and hence the choice of ϵ in Step 0.v directly influences the accuracy of Algorithm 3.22 and the efficiency of design $\xi_\epsilon^\star$. We require from the efficiency that $\frac{\Phi_\alpha(\xi_\epsilon^\star)}{\Phi_\alpha(\xi^\star)} > \mathit{eff}$, where eff is e.g. 0.999. Usually, the value $\Phi_\alpha(\xi^\star)$ is not known, so, after running Algorithm 3.22, one may at least check whether $\frac{\epsilon}{\Phi_\alpha(\xi_\epsilon^\star)} < 1 - \mathrm{eff}$, since $1 - \frac{\Phi_\alpha(\xi_\epsilon^\star)}{\Phi_\alpha(\xi^\star)} = \frac{\Phi_\alpha(\xi^\star) - \Phi_\alpha(\xi_\epsilon^\star)}{\Phi_\alpha(\xi^\star)} < \frac{\epsilon}{\Phi_\alpha(\xi_\epsilon^\star)}$.

In the case when the exact expressions for the expectations in $\varphi(\mathbf{z})$ and $\nabla\varphi(\mathbf{z})$ (Eqs. (3.21) and (3.23)) are not easily computed, but we are able to generate the random realizations from the prior distribution $\pi(\boldsymbol{\theta})$, we may use Monte Carlo methods to obtain estimates of required expected values (similarly as in Valenzuela et al., 2015, Rockafellar and Uryasev, 2000, or Atkinson et al., 2007, Sect. 18.5), see also Example 3.24. However, Monte Carlo simulation does not ensure the validity of (3.28).

3.6 Examples

The examples in this section were carried out in `R` computing environment (R Core Team, 2016), the LP problems were solved with linear and integer programming solver `lp_solve` (Berkelaar et al., 2004) via `R` package `lpSolveAPI` (lp_solve and Konis., 2016). The solver uses revised simplex method for LP problems. The main purpose

of this section is to point out that by applying Algorithm 3.22 we may obtain nearly CVaR-optimal designs. The obtained results are compared with the known AVE, maximin and local optimal designs (see their relations in Theorem 3.19).

Example 3.23. Consider the nonlinear regression model $\eta(x,\theta) = e^{-x\theta}$. The unknown parameter θ has discrete distribution: $\pi(\theta) = 0.2$ for any $\theta \in \Theta = \left\{1/7, 1/\sqrt{7}, 1, \sqrt{7}, 7\right\}$, as in Example 18.2 of Atkinson et al. (2007). We take the discretization $\mathcal{X}' = \{0, 0.001, \ldots, 6.999, 7\}$ and the starting design $\xi^{(0)}$ as the uniform measure on $\mathcal{X}'$. We set $c^{(0)} = 2$ and $\epsilon = 10^{-6}$. In this case, the required expectations can be quite simply computed analytically and hence we do not need to use Monte Carlo methods. The goal is to approach Φ_α-optimal design for $\phi_D(\xi, \boldsymbol{\theta}) = \det{}^{1/m}[M(\xi, \boldsymbol{\theta})]$ via Algorithm 3.22. For nonsingular information matrix $M\left(\widetilde{\xi}, \boldsymbol{\theta}\right)$ one obtains $\left\{\nabla_\xi \phi_D\left(\widetilde{\xi}, \boldsymbol{\theta}\right)\right\}_i = \frac{\det^{1/m}\left[M\left(\widetilde{\xi}, \boldsymbol{\theta}\right)\right]}{m} \frac{\partial \eta(x,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big|_{x=x_i} M^{-1}\left(\widetilde{\xi}, \boldsymbol{\theta}\right) \frac{\partial \eta^\top(x,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{x=x_i}$, the i-th component of gradient of $\phi_D(\xi, \boldsymbol{\theta})$ at $\widetilde{\xi}$ (see e.g. Pronzato and Pázman, 2013, Example 9.17).

The results of Algorithm 3.22 for different levels α are summarized in Table 3.1.

α	$\xi_\epsilon^\star$	$\Phi_\alpha(\xi_\epsilon^\star)$	$c_\epsilon^\star$	iter.	time
1	$\begin{Bmatrix} 6.522 \\ 1 \end{Bmatrix}$	1.3813	27728.31	2	5.36 s
0.5	$\begin{Bmatrix} 0.847 \\ 1 \end{Bmatrix}$	0.0296	0.1318	4	5.43 s
0.3	$\begin{Bmatrix} 0.320 \\ 1 \end{Bmatrix}$	0.0071	0.0188	5	5.85 s
0.22	$\begin{Bmatrix} 0.179 \\ 1 \end{Bmatrix}$	0.0035	0.0124	5	5.74 s
0.2	$\begin{Bmatrix} 0.143 \\ 1 \end{Bmatrix}$	0.0028	0.0096	5	6.04 s
0.0001	$\begin{Bmatrix} 0.143 \\ 1 \end{Bmatrix}$	0.0028	0.0028	5	5.69 s

Table 3.1: The numerical results of Example 3.23. All the computed ϵ-optimal designs are single point designs and hence $\xi_\epsilon^\star$ is a Dirac measure putting unit mass to the optimal design point (second column). In the last two columns is the number of iterations and the computational time (both required until Algorithm 3.22 stopped) indicated.

The ϵ-optimal design which maximizes the criterion $\Phi_1(\xi)$ corresponds to the AVE-optimal design, see Table 18.4 in Atkinson et al. (2007). The value indicated as $c_\epsilon^\star = 27728.31$ equals here $c_{up} = \max_{\theta\in\Theta} \phi_D(\mathbf{1}, \theta)$.

According to the theoretical results on maximin design for this nonlinear regression model (see e.g. Example 3.18 or Pronzato and Pázman, 2013, Example 8.5), the optimal maximin design puts unit mass to the point $\frac{1}{\max_{\theta\in\Theta}\theta}$, which is for our Θ equal to $1/7 \approx 0.1428$ and corresponds to the ϵ-optimal design which maximizes $\Phi_{0.0001}(\xi)$. Since the distribution of $\phi_D(\xi,\theta)$ is uniform on a finite set with cardinality five, the values of $\Phi_\alpha(\xi)$ are equal to $\min_{\theta\in\Theta}\phi_D(\xi,\theta)$ for any $\alpha \in (0, 0.2]$. For $\alpha \in (0.2, 0.4)$ the criteria $\Phi_\alpha(\xi)$ do not attain the same value (although the maximum in (3.7) is reached for the same $c = q_{\phi(\xi,\theta)}(\alpha) = Rq_{\phi(\xi,\theta)}(\alpha)$) because the expectation in (3.7) is multiplied by $\frac{1}{\alpha}$. △

Example 3.24. We consider the nonlinear regression model of Atkinson et al. (1993) with $\eta(x,\boldsymbol{\theta}) = \theta_3\left(e^{-\theta_1 x} - e^{-\theta_2 x}\right)$, $\boldsymbol{\theta} = (\theta_1,\theta_2,\theta_3) \in \Theta = \left[\theta_1^0 - 0.01, \theta_1^0 + 0.01\right] \times$ $\times \left[\theta_2^0 - 1, \theta_2^0 + 1\right] \times \theta_3^0$, where $\boldsymbol{\theta}^0 = \left(\theta_1^0,\theta_2^0,\theta_3^0\right)^\top = (0.05884, 4.298, 21.8)^\top$. The prior distribution π_Θ is uniform over Θ. The aim of this example is to find Φ_α-optimal designs on the design space $\mathcal{X} = [0,\infty)$ via the cutting plane method. If the accurate results are known from Tables 1–2 of Atkinson et al. (1993), we evaluate the efficiencies of designs computed by Algorithm 3.22 with respect to the true optimal designs of Atkinson et al. (1993) (see Table 3.2).

As we know, the criterion $\Phi_\alpha(\xi)$ is not invariant to nonlinear rescaling, and, to be consistent with Atkinson et al. (1993), in this example we use the criterion of D-optimality in the form $\phi_{D*}(\xi,\boldsymbol{\theta}) = \ln\det\left[M(\xi,\boldsymbol{\theta})\right]$. The i-th component of required gradient of $\phi_{D*}(\xi,\boldsymbol{\theta})$ at $\widetilde{\xi}$ is, when $M\left(\widetilde{\xi},\boldsymbol{\theta}\right)$ is nonsingular, $\left\{\nabla_\xi\phi_{D*}\left(\widetilde{\xi},\boldsymbol{\theta}\right)\right\}_i = \frac{\partial\eta(x,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\Big|_{x=x_i} M^{-1}\left(\widetilde{\xi},\boldsymbol{\theta}\right)\frac{\partial\eta^\top(x,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{x=x_i}$ (see Pronzato and Pázman, 2013, Example 9.17).

The first encountered problem is that the expectations in (3.21) and (3.23) are not easily expressed. Fortunately, we can effectively generate random variables from the prior distribution π_Θ and hence the Monte Carlo method is applicable. Let $\boldsymbol{\theta}^1,\ldots\boldsymbol{\theta}^R$ be a random sample from π_Θ generated during the Step 0 of Algorithm 3.22, i.e. the same sample is used for all iterations and for evaluating $c_{up} = \max_{h\in\{1,\ldots,R\}}\phi_{D*}\left(\mathbf{1},\boldsymbol{\theta}^h\right)$. In Algorithm 3.22 we will use the expression $c + \frac{1}{R\alpha}\sum_{h=1}^R \min\left\{0, \phi\left(\xi,\boldsymbol{\theta}^h\right) - c\right\}$ to approach φ in (3.21) and instead of $\nabla\varphi$ in (3.23) we will have

$$\begin{pmatrix}\mathbf{0}\\1\end{pmatrix} + \frac{1}{R\alpha}\sum_{h=1}^R \begin{cases} \begin{pmatrix}\mathbf{0}\\0\end{pmatrix} & \text{if } \phi\left(\widetilde{\xi},\boldsymbol{\theta}^h\right) > \widetilde{c}, \\ \begin{pmatrix}\nabla_\xi\phi\left(\widetilde{\xi},\boldsymbol{\theta}^h\right)\\-1\end{pmatrix} & \text{otherwise} \end{cases}.$$

The second problem is that we are not able to evaluate the gradient of $\phi_{D*}(\xi,\boldsymbol{\theta})$ when $M(\xi,\boldsymbol{\theta})$ does not have full rank. Hence at each iteration when computing

$M\left(\xi^{(j)}, \boldsymbol{\theta}^h\right)$ for $j = 1, \ldots, i$, we perform the following regularization step for any $\boldsymbol{\theta}^h$, $h = 1, \ldots, R$:

$$\text{if } \det\left[M\left(\xi^{(j)}, \boldsymbol{\theta}^h\right)\right] < 10^{-8}, \text{ then } M\left(\xi^{(j)}, \boldsymbol{\theta}^h\right) \leftarrow M\left(\xi^{(j)}, \boldsymbol{\theta}^h\right) + 10^{-5}\mathbf{I}_m.$$

We ran Algorithm 3.22 three times. In the first run the discretization of the design space was $\mathcal{X}' = \{0.2, 0.3, \ldots, 21.9, 22\}$. Suppose that $x_1^\star, \ldots, x_S^\star$ are the support points of the resulting design. In the second run we used the discretization $\cup_{j=1}^{S}\left\{x_j^\star - 0.09, x_j^\star - 0.08, \ldots, x_j^\star + 0.09\right\} \cup \mathcal{X}'$, and similarly in the third run, where we enriched the previous discretization with sets $\{x^\star - 0.009, x^\star - 0.008, \ldots, x^\star + 0.009\}$ for every $x^\star$, the support point obtained after the second run. In all the runs the starting design $\xi^{(0)}$ was chosen randomly (respecting its non-negativity and summation to one). For each run we set $c^{(0)} = c_{up}$ (c_{up} here takes the value around 21 up to 25 depending on the length of vector $\mathbf{1}$, i.e. on the cardinality of the discretized design space $\mathcal{X}'$, and is computed as suggested in Sect. 3.5.1), $\epsilon = 10^{-4}$, and $R = 100$.

In Table 3.2 we present CVaR ϵ-optimal designs for different levels α obtained as a result of three runs of Algorithm 3.22. Although the ϵ-optimal designs $\xi_\epsilon^\star$ were computed by means of Monte Carlo simulation with the sample size $R = 100$, the criterion values $\Phi_\alpha(\xi_\epsilon^\star)$ and the efficiencies $\Phi_\alpha(\xi_\epsilon^\star)/\Phi_\alpha^\star$, where $\Phi_\alpha^\star$ denotes the true optimal value known from the literature (Atkinson et al., 1993, Tables 1–2), were evaluated with different sample size $R' = 5 \times 10^5$. △

In the next example we compare the CVaR- and quantile optimal designs. Since the quantile criterion Φ_α^Q is not concave, and hence its optimization is not straightforward, we think that in applications we could use the CVaR-criterion Φ_α and CVaR-optimal designs as approximations of quantile optimal designs.

Example 3.25. Consider the nonlinear regression model with expectation of observed variable at $x \in \mathbb{R}_0^+$ equal to $\eta(x, \boldsymbol{\theta}) = \theta_1 e^{-\theta_2 x}$, where $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top \in \mathbb{R}^+ \times [0.5; 3.5]$. The example is from Pázman and Pronzato (2007), where the authors considered the D-efficiency criterion

$$\phi_{\text{effD}}(\xi, \boldsymbol{\theta}) = \frac{\sqrt{\det M(\xi, \boldsymbol{\theta})}}{\frac{\theta_1}{2e\theta_2}} \in [0, 1], \tag{3.29}$$

(since $\max_{\xi \in \Xi} \sqrt{\det M(\xi, \boldsymbol{\theta})} = \frac{\theta_1}{2e\theta_2}$, see Pázman and Pronzato, 2007).

The parameter θ_2 has uniform prior on $[0.5; 3.5]$ and we set $\theta_1 = 1$ and $\mathcal{X} = \{0, 0.1, 0.2, \ldots, 5\}$ as the design space consisting of 51 points as in Pázman and Pronzato (2007).

To compute the CVaR-optimal designs we apply Algorithm 3.22 similarly as in Example 3.24, and we use the gradient of D-efficiency criterion $\nabla_\xi \phi_{\text{effD}}\left(\widetilde{\xi}, \boldsymbol{\theta}\right) =$

prior	α	$\xi_\epsilon^\star$	$c_\epsilon^\star$	$\Phi_\alpha(\xi_\epsilon^\star)$	efficiency	time
$\pi\left(\boldsymbol{\theta}^0\right)=1$		$\begin{Bmatrix}0.229 & 1.389 & 18.417\\ 0.333 & 0.334 & 0.335\end{Bmatrix}$	7.3771	7.3887	0.9999	3.42 s
π_Θ	1	$\begin{Bmatrix}0.230 & 1.420 & 18.570\\ 0.332 & 0.333 & 0.335\end{Bmatrix}$	8.1106	7.3758	0.9999	5.44 m
	0.5	$\begin{Bmatrix}0.214 & 1.320 & 17.230\\ 0.331 & 0.334 & 0.334\end{Bmatrix}$	7.4185	7.1215		5.13 m
	0.1	$\begin{Bmatrix}0.199 & 1.220 & 16.680\\ 0.334 & 0.334 & 0.332\end{Bmatrix}$	7.0304	6.8604		4.30 m

Table 3.2: The numerical results of Example 3.24. The first row corresponds to the locally optimal design with nominal parameter value $\boldsymbol{\theta}^0$. The computing time (last column) was 3.42 seconds for all three runs of Algorithm 3.22 together. The computing times for the rest of the table were higher, since at each iteration the estimates of the expectations in (3.21) and (3.23) were calculated. The ϵ-optimal designs $\xi_\epsilon^\star$ for different probability levels α and the optimal $c_\epsilon^\star$ were computed using Algorithm 3.22 with $R=100$ randomly generated vectors from π_Θ. For $\alpha=1$ we computed $\Phi_1(\xi_\epsilon^\star)$ as $\frac{1}{R'}\sum_{h=1}^{R'}\phi_{D*}\left(\xi_\epsilon^\star,\boldsymbol{\theta}^h\right)$, with $R'=5\times10^5$, and the efficiency as $\frac{\Phi_1(\xi_\epsilon^\star)}{\Phi_1^\star}$, where $\Phi_1^\star$ is the optimal value from Table 2 in Atkinson et al. (1993). For $\alpha<1$ the value indicated is $\Phi_\alpha(\xi_\epsilon^\star)=c_\epsilon^\star+\frac{1}{\alpha R'}\sum_{h=1}^{R'}\left[\min\left\{0,\phi_{D*}\left(\xi,\boldsymbol{\theta}^h\right)-c_\epsilon^\star\right\}\right]$, with $R'=5\times10^5$.

$\frac{2e\theta_2}{\theta_1}\nabla_\xi\phi_D\left(\widetilde{\xi},\boldsymbol{\theta}\right)$ (see Example 3.23 for $\nabla_\xi\phi_D\left(\widetilde{\xi},\boldsymbol{\theta}\right)$). Instead of generating R realizations from the prior distribution of θ_2, we took only the sequence of $R=100$ points equally spaced in the interval $[0.5;3.5]$, and the expectations in (3.21) and (3.23) were approximated by arithmetic means based on these values. We set the accuracy in the stopping rule as $\epsilon=10^{-3}$ and the initial design $\xi^{(0)}$ is the uniform measure on the design space $\mathcal{X}$. Since the criterion in (3.29) takes the values from $[0,1]$, we set $c_{low}=0$ and $c^{(0)}=c_{up}=1$. The CVaR ϵ-optimal designs are for $\alpha=0.1$ and $\alpha=0.5$ given in Table 3.3.

To compute the quantile optimal designs we proceed as in Pázman and Pronzato (2007): we approximate the quantile criterion and its directional derivatives via kernel smoothing and then we use vertex direction algorithm (see e.g. Pronzato and Pázman, 2013, Sect. 9.1.1) to approach the quantile optimal designs for $\alpha=0.1$ and for $\alpha=0.5$. Denote by δ_x the design with the unit mass at the design point x and let $\widetilde{\mathcal{F}_{\Phi_\alpha^Q}}(\xi,\zeta)$ be the approximation of the directional derivative of Φ_α^Q at the point ξ in the direction ζ. The vertex direction algorithm, which we used, consisted of these steps:

Algorithm 3.26.

0. Choose the starting design $\xi^{(0)}$ and the accuracy $\epsilon>0$. Set $i=0$. Find $x^{(0)}=$

α	$\xi_\epsilon^\star$	$c_\epsilon^\star$	$\Phi_\alpha^{Q_{10000}}(\xi_\epsilon^\star)$	time
0.1	$\begin{Bmatrix} 0 & 0.3 & 0.4 & 1.6 \\ 0.452 & 0.148 & 0.210 & 0.191 \end{Bmatrix}$	0.7803	0.7807	5.50 s
0.5	$\begin{Bmatrix} 0 & 0.5 & 1.1 \\ 0.501 & 0.478 & 0.020 \end{Bmatrix}$	0.9246	0.9231	4.11 s

Table 3.3: The numerical results of Example 3.25. The CVaR ϵ-optimal designs $\xi_\epsilon^\star$ for the D-efficiency criterion (3.29) for probability level α and the optimal $c_\epsilon^\star$ computed by Algorithm 3.22 based on $R = 100$ equally spaced θ_2 in the interval $[0.5; 3.5]$. The value $\Phi_\alpha^{Q_{10000}}(\xi_\epsilon^\star)$ is the direct approximation of corresponding quantile criterion value based on 10000 equally spaced θ_2 in the interval $[0.5; 3.5]$. The last column gives the running time of Algorithm 3.22.

$\arg\max_{x\in\mathcal{X}} \widetilde{\mathcal{F}_{\Phi_\alpha^Q}}\left(\xi^{(0)}, \delta_x\right)$.

1. Set $\xi^{(i+1)} = \left(1 - \frac{1}{2+i}\right)\xi^{(i)} + \frac{1}{2+i}\delta_{x^{(i)}}$.
2. Set $i \leftarrow i+1$.
3. Find $x^{(i)} = \arg\max_{x\in\mathcal{X}} \widetilde{\mathcal{F}_{\Phi_\alpha^Q}}\left(\xi^{(i)}, \delta_x\right)$.
4. If $\widetilde{\mathcal{F}_{\Phi_\alpha^Q}}\left(\xi^{(i)}, \delta_{x^{(i)}}\right) > \epsilon$ continue from Step 1, else take $\xi^{(i)}$ as ϵ-optimal design and stop.

We would like to emphasize that the purpose of this example and of this thesis is not to find the effective methodology for computing quantile optimal designs and there may be other more effective procedures how to obtain the best solution. Nevertheless, here we took $\epsilon = 0.01$ and the kernel-approximation of the quantile criterion and of the directional derivative as in Pázman and Pronzato (2007):

$$\widetilde{\Phi_\alpha^Q}(\xi) = \left\{ u : 1 - \frac{1}{R}\sum_{i=1}^{R} F_{\mathcal{N}(0,1)}\left(\frac{u - \phi_{\text{effD}}\left(\xi, \boldsymbol{\theta}^i\right)}{h_R(\xi)}\right) = 1 - \alpha \right\},$$

$$\widetilde{\mathcal{F}_{\Phi_\alpha^Q}}(\xi, \zeta) = \frac{\sum_{i=1}^{R} \mathcal{F}_{\phi_{\text{effD}}(\cdot,\boldsymbol{\theta}^i)}(\xi, \zeta)\, f_{\mathcal{N}(0,1)}\left(\frac{\widetilde{\Phi_\alpha^Q}(\xi) - \phi_{\text{effD}}\left(\xi,\boldsymbol{\theta}^i\right)}{h_R(\xi)}\right)}{\sum_{i=1}^{R} f_{\mathcal{N}(0,1)}\left(\frac{\widetilde{\Phi_\alpha^Q}(\xi) - \phi_{\text{effD}}\left(\xi,\boldsymbol{\theta}^i\right)}{h_R(\xi)}\right)},$$

where $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^R$ is a set of vector points with first coordinates equal to 1 (corresponding to θ_1) and with second coordinates given as a sequence of R equally spaced points in $[0.5; 3.5]$. In accordance with Pázman and Pronzato (2007), we used $R = 100$ and we selected $h_R(\xi) = s_R(\xi)\, R^{-1/5}$ with $s_R(\xi) = \sqrt{\frac{\sum_{i=1}^{R}\left[\phi_{\text{effD}}\left(\xi,\boldsymbol{\theta}^i\right) - \frac{1}{R}\sum_{j=1}^{R}\phi_{\text{effD}}\left(\xi,\boldsymbol{\theta}^j\right)\right]^2}{R-1}}$,

the standard deviation of $\phi_{\text{effD}}\left(\xi,\boldsymbol{\theta}^1\right),\ldots,\phi_{\text{effD}}\left(\xi,\boldsymbol{\theta}^R\right)$. The functions $F_{\mathcal{N}(0,1)}(\cdot)$ and $f_{\mathcal{N}(0,1)}(\cdot)$ denote the cdf and pdf of the standardized normal random variable $\mathcal{N}(0,1)$ and $\mathcal{F}_{\phi_{\text{effD}}(\cdot,\boldsymbol{\theta})}(\xi,\zeta)=\frac{e\theta_2}{\theta_1}\det{}^{1/2}\left[M(\xi,\boldsymbol{\theta})\right]\text{tr}\left\{\left[M(\zeta,\boldsymbol{\theta})-M(\xi,\boldsymbol{\theta})\right]M^{-1}(\xi,\boldsymbol{\theta})\right\}$ is the directional derivative of $\phi_{\text{effD}}(\cdot,\boldsymbol{\theta})$ at ξ in the direction ζ.

As mentioned in Pázman and Pronzato (2007) (and as easily computed via Algorithm 3.22 when putting Dirac measure concentrated at $\theta_2=2$ as prior for θ_2), $\xi_{\text{loc}}=\begin{Bmatrix}0 & 1/2\\ 1/2 & 1/2\end{Bmatrix}$ is the locally D-optimal design for $\theta_2=2$.

As a starting design in Algorithm 3.26 we took the CVaR-optimal design for given probability level α (see Table 3.3), the local optimal design ξ_{loc} for $\theta_2=2$ (as indicated above) and the design which puts the same mass $1/\text{card}(\mathcal{X})$ to each design point x from $\mathcal{X}$, respectively. The results are summarized in Table 3.4. Comparing Tables 3.3 and 3.4 one sees that the CVaR and quantile optimal values do not differ significantly, which indicates the possibility of using the CVaR-criterion instead of the non-concave quantile criterion. Indeed, the Φ_α^Q-efficiency of CVaR-optimal design equals 0.7807/0.7930=0.9845 for $\alpha=0.1$ and 0.9231/0.9495=0.9721 for $\alpha=0.5$.

The histograms of $\phi_{\text{effD}}(\xi,\boldsymbol{\theta})$, where θ_2 is distributed according to its prior distribution (uniform on $[0.5;3.5]$), are for some designs ξ given in Figure 3.6. While the quantile optimal design ensures the largeness of the α right-quantile (without taking into account the smallness of values less than the right-quantile), the CVaR-criterion maximizes the expectation of values less than the corresponding α right-quantile.

The performances of the quantile, AVE, maximin and CVaR-optimal designs are illustrated in Figure 3.7, where the corresponding values of the D-efficiency criterion ϕ_{effD} are depicted as functions of θ_2 (we obtained the AVE and maximin optimal designs as Φ_α-optimal designs setting $\alpha=1$ resp. $\alpha=0.0001$, see Theorem 3.19 for a justification). The totally robust design which does not depend on the true parameter value θ_2 would be in this figure represented by a horizontal line. For $\alpha=0.1$, the AVE-optimal design gives larger values of ϕ_{effD} than the quantile, maximin and CVaR optimal designs for a wide range of possible values of θ_2. On the other hand, it performs very bad for θ_2 near 0.5, where the best results are given by the maximin optimal design. The quantile and CVaR-optimal designs behave similarly, as a compromise between the AVE and maximin optimal design—the CVaR-criterion performs slightly better near 0.5, and the quantile criterion is better for the rest of the values. For $\alpha=0.5$, the CVaR and quantile (median) optimal designs behave very similar to the AVE optimal design, the quantile optimal design attains better values for larger θ_2 and the CVaR-optimal design for the smaller values of θ_2, where the performance of the quantile optimal design is very poor.

Finally, a very interesting interpretation follows from the comparison of empirical cdfs of $\phi_{\text{effD}}(\xi, \boldsymbol{\theta})$, where ξ is AVE, maximin, CVaR, or quantile optimal design. The graphs of cdfs in $(0,1)$ are displayed in Figure 3.8. The criterion $\phi_{\text{effD}}(\xi, \boldsymbol{\theta})$ always attains values from the interval $[0,1]$ and the "perfect design" would lead to the cdf $F_{\phi_{\text{effD}}(\cdot,\boldsymbol{\theta})}(t)$ which equals zero for $t < 1$ and equals one for $t \geq 1$. Obviously, this is not very realistic.

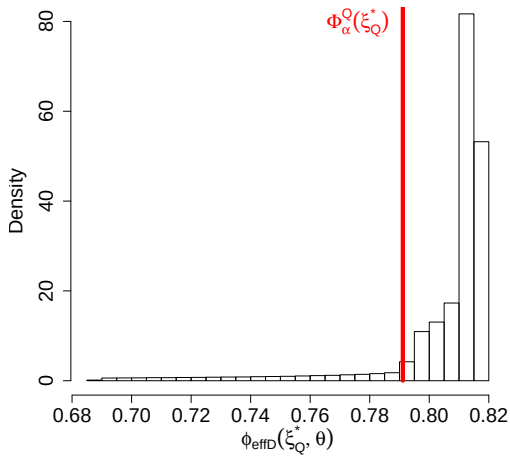
In Figure 3.8a is the situation for $\alpha = 0.1$ depicted. One sees that for $t = 0.775$, the probability $Pr[\phi_{\text{effD}}(\xi^{\star}_{\text{min}}, \boldsymbol{\theta}) \leq t] = 0.32$, while $Pr[\phi_{\text{effD}}(\xi^{\star}_{\text{CVaR}}, \boldsymbol{\theta}) \leq t] = 0.04$. Hence, by considering the CVaR optimal design and "sacrificing" the values $\phi_{\text{effD}}(\xi^{\star}_{\text{CVaR}}, \boldsymbol{\theta})$ which are smaller than $\min_{\boldsymbol{\theta} \in \Theta} \phi_{\text{effD}}(\xi^{*}_{\text{min}}, \boldsymbol{\theta})$, we obtain significant improvement at $t = 0.775$ comparing to the maximin optimal design. The approach via AVE criterion is not robust, and the AVE-optimal design leads with nonzero probability to criterion values smaller than 0.6.

In Figure 3.8b is indicated how could we approach the cdf corresponding to 10%-quantile optimal design by considering the cdfs of 20%- and 25%-CVaR optimal designs. It would be very interesting if we had some theoretical results related to the possibilities of approaching the quantile optimal design via optimizing the CVaR criterion.

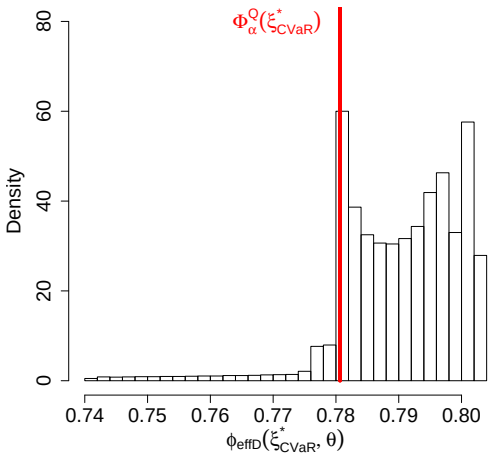
Figure 3.8c displays the cdfs for $\alpha = 0.5$. △

α	$\xi^{(0)}$	$\xi_\epsilon^\star$	$\Phi_\alpha^{Q_{kernel}}(\xi_\epsilon^\star)$	$\Phi_\alpha^{Q_{10000}}(\xi_\epsilon^\star)$	time
0.1	CVaR	$\begin{Bmatrix} 0 & 0.3 & 0.4 & 0.6 & 1.2 & 1.3 & 1.4 & 1.6 \\ 0.459 & 0.253 & 0.097 & 0.007 & 0.048 & 0.129 & 0.007 & 0.001 \end{Bmatrix}$	0.7851	0.7924	**1320** s
	local	$\begin{Bmatrix} 0 & 0.3 & 0.4 & 0.5 & 1.1 & 1.2 & 1.3 & 1.4 \\ 0.457 & 0.246 & 0.107 & 0.003 & 0.005 & 0.043 & 0.134 & 0.005 \end{Bmatrix}$	0.7853	**0.7930**	1642 s
	uniform	$\begin{Bmatrix} 0 & 0.3 & 0.4 & 0.5 & 0.9 & 1.2 & 1.3 & 1.4 \\ 0.459 & 0.248 & 0.096 & 0.009 & 0.005 & 0.018 & 0.156 & 0.005 \end{Bmatrix}$	0.7839	0.7911	1910 s
0.5	CVaR	$\begin{Bmatrix} 0 & 0.4 & 0.5 & 1.1 \\ 0.500 & 0.429 & 0.068 & 0.003 \end{Bmatrix}$	0.9305	**0.9495**	60.91 s
	local	$\begin{Bmatrix} 0 & 0.4 & 0.5 \\ 0.500 & 0.400 & 0.100 \end{Bmatrix}$	0.9315	0.9493	**43.18** s
	uniform	$\begin{Bmatrix} 0 & 0.4 & 0.5 & 0.6 \\ 0.495 & 0.457 & 0.029 & 0.010 \end{Bmatrix}$	0.9251	0.9449	904 s

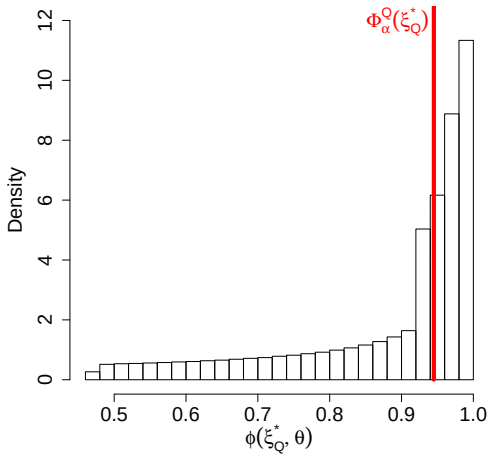
Table 3.4: The numerical results of Example 3.25. The ϵ-optimal designs $\xi_\epsilon^\star$ for the quantile criterion Φ_α^Q with probability level α computed via the vertex direction algorithm 3.26 initialized at the design $\xi^{(0)}$. The value $\Phi_\alpha^{Q_{kernel}}(\xi_\epsilon^\star)$ is the approximation of corresponding α-quantile criterion computed using Kernel smoothing with $R = 100$ equally spaced θ_2 in the interval $[0.5; 3.5]$ and the value $\Phi_\alpha^{Q_{10000}}(\xi_\epsilon^\star)$ is the direct approximation of corresponding α-quantile criterion based on 10000 equally spaced θ_2 in the interval $[0.5; 3.5]$. The last column gives the running time of Algorithm 3.26.



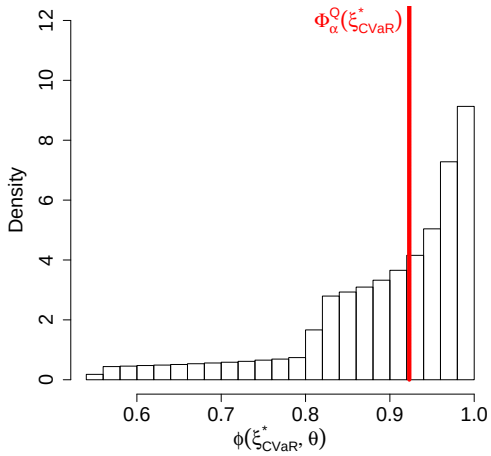
(a) Φ_α^Q-optimal design, $\alpha = 0.1$



(b) Φ_α-optimal design, $\alpha = 0.1$

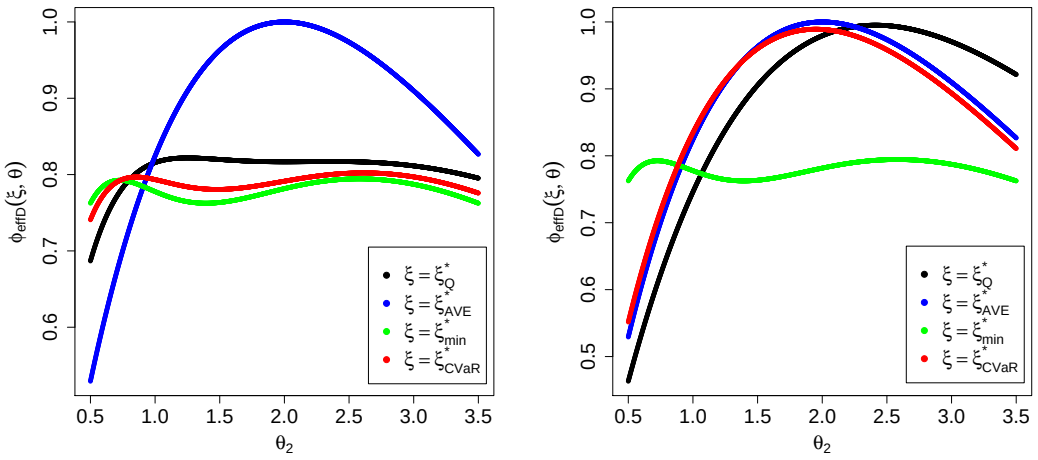


(c) Φ_α^Q-optimal design, $\alpha = 0.5$



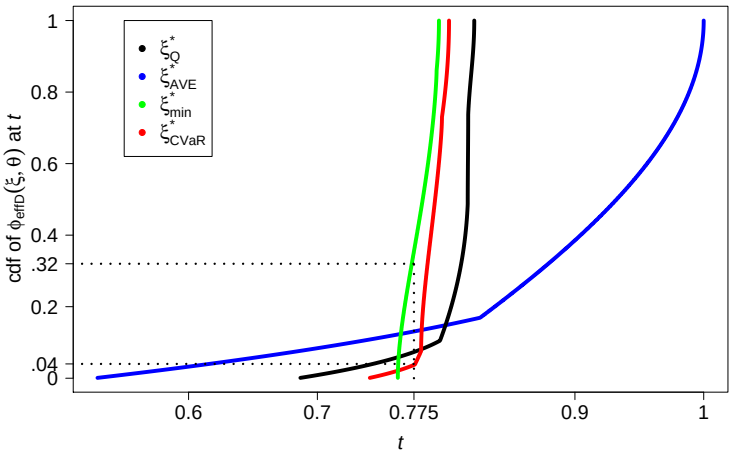
(d) Φ_α-optimal design, $\alpha = 0.5$

Figure 3.6: Example 3.25: The histograms of $\phi_{\text{effD}}(\xi, \boldsymbol{\theta})$ based on 10000 equally spaced θ_2 in the interval $[0.5; 3.5]$, where ξ is either Φ_α^Q or Φ_α-optimal design for $\alpha = 0.1, 0.5$. In each subfigure is the value $\Phi_\alpha^Q(\xi)$ depicted.

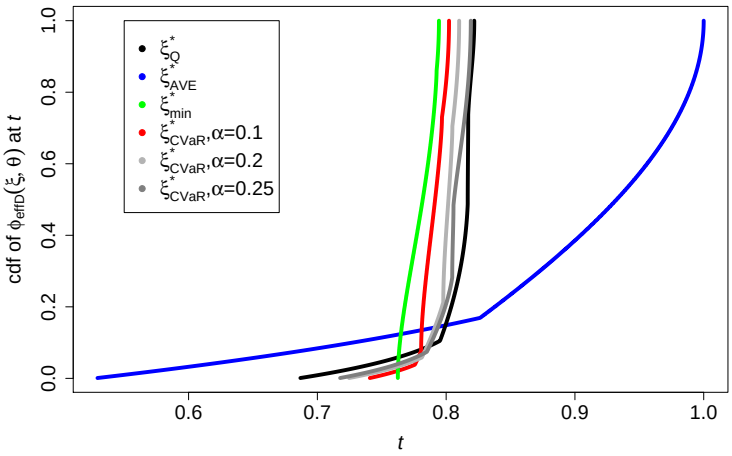


(a) $\alpha = 0.1$ (b) $\alpha = 0.5$

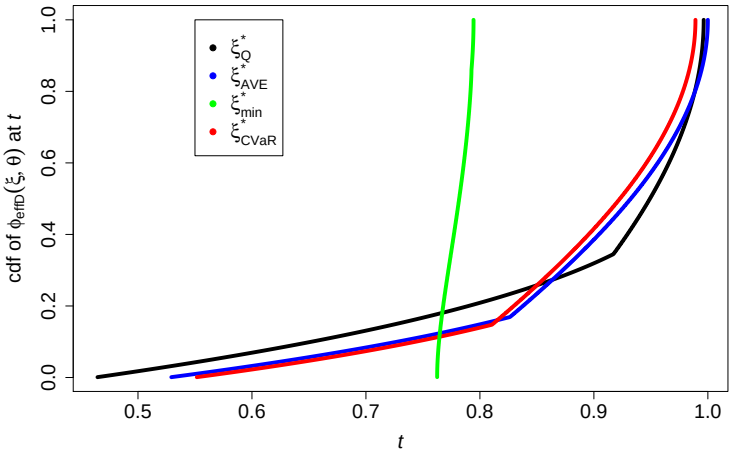
Figure 3.7: Example 3.25: Criterion $\phi_{\text{effD}}(\xi, \boldsymbol{\theta})$ as a function of θ_2 for quantile Φ^Q_α, AVE Φ_1, maximin $\Phi_{0.0001}$ and CVaR Φ_α-optimal designs.



(a) $\alpha = 0.1$



(b) $\alpha = 0.1$



(c) $\alpha = 0.5$

Figure 3.8: Example 3.25: empirical cdfs of $\phi_{\text{effD}}(\xi, \boldsymbol{\theta})$ beased on 10000 equally spaced θ_2 in the interval $[0.5; 3.5]$, where ξ is maximin $\Phi_0.0001$, AVE Φ_1, CVaR, or quantile optimal design.

3.7 Conclusions to this chapter

The Conditional Value at Risk was used for the design purposes (input design problem) in Valenzuela et al. (2015) for the first time. We think that the introduction of the CVaR criterion in experimental design is very interesting, since it enlarges the possibilities of the popular AVE and maximin criteria—which are not quite satisfactory from a certain point of view—and is related to the recently proposed quantile criterion (Pázman and Pronzato, 2007), which is not concave.

In Sects. 3.2–3.3 we summarized and explained the properties of the CVaR-criterion. We were inspired by results obtained in the risk theory (Pflug, 2000; Rockafellar and Uryasev, 2000, 2002). The formula for the CVaR-criterion used in this thesis differs from the formula introduced in Valenzuela et al. (2015). Its concavity is ensured (when $\phi(\xi, \boldsymbol{\theta})$ is a concave function of ξ) whether the random variable $\phi(\xi, \boldsymbol{\theta})$ is continuous or not. We think that the expansion to non-continuous random variables is important because of the possibility of finiteness of the parametric space Θ, which necessary leads to a discrete random variable $\phi(\xi, \boldsymbol{\theta})$. This motivated our detailed investigations in Sects. 3.2–3.3, since the case of non-continuous random variable $\phi(\xi, \boldsymbol{\theta})$ turned out to be more difficult than the continuous one.

We showed that the CVaR-criterion $\Phi_\alpha(\xi)$ can be considered as a compromise between AVE ($\alpha = 1$) and maximin ($\alpha \to 0$) criterion, see Sect. 3.4, which gave us one possible interpretation of $\Phi_\alpha(\xi)$. As mentioned in Pázman and Pronzato (2007), the AVE-optimal design $\xi^\star_{\text{AVE}}$ does not exclude the situation that the probability $Pr[\phi(\xi^\star_{\text{AVE}}, \boldsymbol{\theta}) \leq t]$ is large for given small t. This drawback is partially treated by CVaR-criterion, which is related to the expected value of the criterion $\phi(\xi, \boldsymbol{\theta})$ under the condition that $\phi(\xi, \boldsymbol{\theta})$ is less than (or less than or equal to) the corresponding α-right-quantile $Rq_{\phi(\xi,\boldsymbol{\theta})}(\alpha)$, see Sect. 3.3. Pázman and Pronzato (2007) pointed out that the maximin optimal design is often focused on $\boldsymbol{\theta}$ at the boundary of the parametric space Θ, which is not the case of AVE, CVaR or quantile criteria as illustrated in Example 3.18.

Under some assumptions, we were able to formulate the equivalence theorem for the CVaR criterion, and we derived the subgradient and the directional derivative of the function $w_\alpha(\xi, c)$ related to the CVaR optimality criterion. We suggested a methodology how to obtain CVaR-optimal designs when $\phi(\xi, \boldsymbol{\theta})$ is concave in ξ via the method of cutting planes in Sect. 3.5, which was illustrated on examples in Sect. 3.6.

In the future, we could investigate the proper choice for probability level α

1. to obtain design which is not focused only on some values $\boldsymbol{\theta}$ like maximin optimal design but which is still sufficiently robust or
2. to approximate the quantile optimal design as indicated at the end of Example 3.25.

Finally, we would like to focus on application of Jeffrey's non-informative prior when there is no known prior on the parametric space Θ.

Chapter 4

Extended optimality criteria for avoiding false estimates in generalized regression models

In this chapter we deal with possible instabilities related to the maximum likelihood estimation in generalized regression models. The aim is to prevent the problems with the uniqueness and identifiability of the maximum likelihood estimate at the design stage of the experiment.

To achieve this goal, we follow Pázman and Pronzato (2014), who introduced the extended criteria of E-, **c**-, and G-optimality to avoid these problems with least squares estimation in classical nonlinear regression. For deeper analysis we also refer the reader to Chap. 7 of the monograph Pronzato and Pázman (2013). We were able to extend the results of Pázman and Pronzato (2014) to generalized regression models based on exponential families of distributions, and, in addition, we also derive the extensions of some other optimality criteria, e.g. the MV- and A-optimality criteria. Partial results from this chapter were published in Burclová and Pázman (2016b).

The chapter requires a scrupulous introduction not only to exponential families of distributions and generalized regression models based on them, but also to matrix norms and pseudonorms. We provide such an introduction in Sect. 4.1. In Sect. 4.2 we show how to design experiments in generalized regression models (even by applying the results of previous chapters of this thesis). In Sect. 4.3 we indicate the stability problems related to the maximum likelihood estimation and we introduce the extended optimality criterion based on the I-divergence. In Sects. 4.4–4.5 we define the extensions for some classical optimality criteria. Section 4.6 summarizes some properties of the extended optimality criteria and Sect. 4.7 shows a method based on Pázman and Pronzato (2014) how to calculate the optimal designs for them. Finally, Sect. 4.8

concludes.

4.1 Introduction

4.1.1 Matrix Norms and Pseudonorms

Definition 4.1. The matrix norm is a mapping $\|\cdot\| : \mathbb{R}^{m\times k} \mapsto [0,\infty)$, $A \to \|A\|$ such that $\forall A, B \in \mathbb{R}^{m\times k}$:

a) $\|A\| \geq 0 \quad \wedge \quad [\|A\| = 0 \Leftrightarrow A = \mathbf{0}]$,

b) $\|aA\| = |a|\,\|A\| \quad \forall a \in \mathbb{R}$,

c) $\|A+B\| \leq \|A\| + \|B\|$.

See e.g. Golub and Van Loan (1996), Sect. 2.3.1. In the literature (e.g. Gentle, 2007, p. 128) is often the consistency property $\|AB\| \leq \|A\|\,\|B\|$ required, however, this is not our case and the properties a)-c) are sufficient.

One sees that for $k = 1$ the mapping $\|\cdot\|$ introduced in Def. 4.1 reduces to a standard vector norm on $\mathbb{R}^m$.

In this thesis we pay attention especially to the vector $\ell(p)$ norm and to the Schatten p norm on square matrices in $\mathbb{R}^{m\times m}$:

$$\forall\, p \geq 1 \quad \forall \mathbf{v} = (v_1,\ldots,v_m) \in \mathbb{R}^m \quad \|\mathbf{v}\|_{\ell(p)} = \left(\sum_{i=1}^{m} |v_i|^p\right)^{1/p} \quad (\ell(p) \text{ norm}),$$

$$\forall\, p \geq 1 \quad \forall A \in \mathbb{R}^{m\times m} \quad \|A\|_{\mathrm{S}(p)} = \left\{\mathrm{tr}\left[\left(AA^\top\right)^{p/2}\right]\right\}^{1/p} \quad (\text{Schatten } p \text{ norm}). \tag{4.1}$$

The Schatten p norm is closely related to the $\ell(p)$ norm since $\|A\|_{\mathrm{S}(p)} = \left\|(s_1(A),\ldots,s_m(A))^\top\right\|_{\ell(p)}$, where $s_i(A) = \sqrt{\lambda_i\left(A^\top A\right)}$ (see e.g. Bhatia, 1997, Eq. IV.31).

Besides the well known Euclidean norm ($\ell(2)$ norm), we will often use its generalization on matrix spaces—Frobenius norm (see e.g. Bhatia, 1997, p. 7, Gentle, 2007, p. 132):

$$\forall A \in \mathbb{R}^{m\times k} \quad \|A\|_F = \sqrt{\sum_{j=1}^{k} \|A_{\cdot j}\|^2_{\ell(2)}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{k} \{A\}^2_{ij}}, \tag{4.2}$$

where $A_{\cdot 1},\ldots,A_{\cdot k} \in \mathbb{R}^m$ denote the columns of the matrix $A \in \mathbb{R}^{m\times k}$ and $\{A\}_{ij}$ is the element on i-th row and j-th column of A. The Frobenius norm on square matrices coincides with the Schatten $p = 2$ norm.

In this thesis we do not restrict ourselves only to norms and hence the definition of matrix pseudonorms is now in place.

Definition 4.2. The matrix pseudonorm is a mapping $|||\cdot|||_P : \mathbb{R}^{m\times k} \mapsto [0,\infty)$, $A \to |||A|||_P$ such that $\forall A, B \in \mathbb{R}^{m\times k}$:

a) $|||A|||_P \geq 0$,

b) $|||aA|||_P = |a|\,|||A|||_P \quad \forall a \in \mathbb{R}$

c) $|||A+B|||_P \leq |||A|||_P + |||B|||_P$.

Obviously, the case $k = 1$ corresponds to a vector pseudonorm. The property b) from Def. 4.2 implies that $|||\mathbf{0}|||_P = 0$. Notice that it follows from Def. 4.1 that every norm fulfils also all properties of a pseudonorm.

The following lemma is a partial generalization of the statement about the equivalence of norms (see e.g. Gentle, 2007, Eq. 3.236)—the well known property of norms in finite-dimensional spaces.

Lemma 4.3. *Let $|||\cdot|||_P$ be a pseudonorm and $\|\cdot\|_F$ be the Frobenius norm in $\mathbb{R}^{m\times k}$. Then there exists a real number $a > 0$ such that $\forall A \in \mathbb{R}^{m\times k} : a|||A|||_P \leq \|A\|_F$.*

Proof. We postponed similarly as Morrow (2013) in his proof of equivalence of vector norms.

The statement of the lemma holds trivially for any matrix A such that $|||A|||_P = 0$ and in such case $a > 0$ can be chosen arbitrarily.

Now, let $A \in \mathbb{R}^{m\times k}$ be a matrix with a positive pseudonorm, i.e. $|||A|||_P \neq 0$. Denote by $E_{ij} \in \mathbb{R}^{m\times k}$ the matrix with one unit entry in the i-th row and j-th column and zeros elsewhere. Then $A = \sum_{i=1}^{m}\sum_{j=1}^{k} \{A\}_{ij} E_{ij}$, where $\{A\}_{ij}$ is the entry in i-th row and j-th column of A and hence

$$
\begin{aligned}
0 < |||A|||_P &\leq \sum_{i=1}^{m}\sum_{j=1}^{k} \left|\{A\}_{ij}\right| |||E_{ij}|||_P \leq \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{k} \left|\{A\}_{ij}\right|^2} \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{k} |||E_{ij}|||_P^2} \\
&= \|A\|_F \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{k} |||E_{ij}|||_P^2},
\end{aligned}
$$

where we used the properties b) and c) from Def. 4.2, Cauchy-Schwarz inequality and the definition of the Frobenius norm (4.2). Emphasizing that $\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{k} |||E_{ij}|||_P^2}$ is positive, we denote $a = \frac{1}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{k} |||E_{ij}|||_P^2}}$ and we have that $a|||A|||_P \leq \|A\|_F$. □

Definition 4.4. The dual norm $\|\cdot\|^D$ of the norm $\|\cdot\|$ on $\mathbb{R}^{m\times k}$ is for any $A \in \mathbb{R}^{m\times k}$ given as the mapping

$$
A \mapsto \|A\|^D = \max_{B\in\mathbb{R}^{m\times k}:\|B\|=1} \left|\mathrm{tr}\left(A^\top B\right)\right|.
$$

See e.g. Eq. IV.50 in Bhatia (1997). Note that the dual norm also satisfies the properties from Def. 4.1, i.e. the dual norm is a norm.

Applying the well known Hölder's inequality for vector variables one obtains that the $\ell(q)$ norm with q satisfying the equation $1 = 1/p + 1/q$ is the dual of the $\ell(p)$ norm. Analogical outcome follows for the Schatten p-norm on square matrices (see e.g. Bhatia, 1997, Proposition IV.2.11 and Exercise IV.2.12 (ii)):

$$\forall A \in \mathbb{R}^{m\times m} : \|A\|^D_{\mathrm{S}(p)} = \|A\|_{\mathrm{S}(q)}, \text{ where } q = \frac{p}{p-1}.$$

One sees that in the case of the $\ell(p)$ and Schatten p norm the dual norm of the dual norm is the initial norm, i.e. $\|\mathbf{v}\|^{DD}_{\ell(p)} = \|\mathbf{v}\|_{\ell(p)}$ and $\|A\|^{DD}_{\mathrm{S}(p)} = \|A\|_{\mathrm{S}(p)}$. Using this and (1.2) one obtains for any symmetric positive definite matrix M that (Dette et al., 1995, Lemma 2.2)

$$\begin{aligned}
\max_{B\in\mathbb{R}^{m\times m}:\|B\|^D_{\mathrm{S}(p)}=1} \mathrm{tr}\left[B^\top M^{-1} B\right] &= \max_{B\in\mathbb{R}^{m\times m}:\|B\|^D_{\mathrm{S}(p)}=1} \max_{A\in\mathbb{R}^{m\times m}:A\neq\mathbf{0}} \frac{\mathrm{tr}^2\left(A^\top B\right)}{\mathrm{tr}\left(A^\top M A\right)} \\
&= \max_{A\in\mathbb{R}^{m\times m}:A\neq\mathbf{0}} \frac{\max_{B\in\mathbb{R}^{m\times m}:\|B\|^D_{\mathrm{S}(p)}=1} \mathrm{tr}^2\left(A^\top B\right)}{\mathrm{tr}\left(A^\top M A\right)} \\
&= \max_{A\in\mathbb{R}^{m\times m}:A\neq\mathbf{0}} \frac{\|A\|^2_{\mathrm{S}(p)}}{\mathrm{tr}\left(A^\top M A\right)} \\
&= \frac{1}{\min_{A\in\mathbb{R}^{m\times m}:\|A\|_{\mathrm{S}(p)}=1} \mathrm{tr}\left(A^\top M A\right)}.
\end{aligned} \tag{4.3}$$

Analogically,

$$\max_{\mathbf{u}\in\mathbb{R}^m:\|\mathbf{u}\|^D_{\ell(p)}=1} \mathbf{u}^\top M^{-1}\mathbf{u} = \frac{1}{\min_{\mathbf{u}\in\mathbb{R}^m:\|\mathbf{u}\|_{\ell(p)}=1} \mathbf{u}^\top M \mathbf{u}}. \tag{4.4}$$

4.1.2 Exponential families of distributions

Focusing on the purposes of the thesis, in this section we will mention the most important properties of exponential families of distributions. For more results we refer the reader to Barndorff-Nielsen (1978), Efron (1978), and Brown (1986).

Denote by $\mathcal{Y} \subseteq \mathbb{R}^l$ a sample space of an experiment, i.e. the space of all possible realizations of a random vector $\mathbf{Y}$.

Definition 4.5. The distribution of the random vector $\mathbf{Y}$ is an exponential family if its probability density function (pdf) or (for discrete random variables) its probability mass function (pmf) f, with respect to the σ-finite measure τ, can be expressed as

$$f(\mathbf{y}, \boldsymbol{\gamma}) = \exp\left\{-\psi(\mathbf{y}) + t^\top(\mathbf{y})\boldsymbol{\gamma} - \kappa(\boldsymbol{\gamma})\right\}, \tag{4.5}$$

where $\boldsymbol{\gamma} \in \Gamma \subseteq \mathbb{R}^r$ is *the canonical parameter*, $t : \mathcal{Y} \to \mathbb{R}^r$, $\psi : \mathcal{Y} \to \mathbb{R}$, and $\kappa : \Gamma \to \mathbb{R}$ are known functions.

It follows that $\mathbf{Y}$ is a continuous random vector in the case of Lebesgue measure τ, and $\mathbf{Y}$ is a discrete random vector if τ is counting measure ($\tau(\{\mathbf{y}\}) = 1$).

Using that $\int_{\mathcal{Y}} f(\mathbf{y}, \boldsymbol{\gamma}) d\tau(\mathbf{y}) = 1$, one can equivalently write $\exp\{\kappa(\boldsymbol{\gamma})\} = \int_{\mathcal{Y}} \exp\left\{-\psi(\mathbf{y}) + t^\top(\mathbf{y})\boldsymbol{\gamma}\right\} d\tau(\mathbf{y})$. Later it will be appropriate to consider the extended space $\Gamma_{\max}$:

$$\Gamma_{\max} = \left\{\boldsymbol{\gamma} \in \mathbb{R}^r : \int_{\mathcal{Y}} \exp\left\{-\psi(\mathbf{y}) + t^\top(\mathbf{y})\boldsymbol{\gamma}\right\} d\tau(\mathbf{y}) < \infty\right\} \supseteq \Gamma. \tag{4.6}$$

Although the canonical parameter $\boldsymbol{\gamma}$ is very useful in the theoretical considerations, it does not necessarily coincide with *the usual parametrization* of given distribution. For instance, consider the binomial distribution $Bin(n, p)$ parametrized by $p \in (0, 1)$ with given n. To obtain the representation of pmf as in (4.5), one has to put $\boldsymbol{\gamma} = \gamma(p) = \ln\left(\frac{p}{p-1}\right) \in \Gamma$ with $\Gamma = \left\{\ln\left(\frac{p}{p-1}\right), p \in (0, 1)\right\}$, see Appendix B.1. It follows that $\boldsymbol{\gamma}$ can be parametrized by some vector parameter $\boldsymbol{\vartheta}$, which may (and may not) correspond to the usual parametrization of the given distribution. Then the density in (4.5) can be equivalently rewritten in the form

$$f[\mathbf{y}, \gamma(\boldsymbol{\vartheta})] = \exp\left\{-\psi(\mathbf{y}) + t^\top(\mathbf{y})\gamma(\boldsymbol{\vartheta}) - \kappa[\gamma(\boldsymbol{\vartheta})]\right\}. \tag{4.7}$$

According to the factorization theorem (see e.g. Billingsley, 1995, Theorem 34.6), $t(\mathbf{Y})$ is a *sufficient statistic*. This allows us to observe $t(\mathbf{Y})$ instead of $\mathbf{Y}$ without losing any information on parameter $\boldsymbol{\gamma}$ (or $\boldsymbol{\vartheta}$), and, moreover, the family of distributions induced by the random variable $t(\mathbf{Y})$ is again an exponential family (Brown, 1986, Proposition 1.5).

Moments in exponential families

Theorem 4.6 (Brown, 1986, Theorem 2.2)**.** *The function* $e^{\kappa(\boldsymbol{\gamma})}$ *is infinitely many times differentiable with respect to* $\boldsymbol{\gamma}$ *at any* $\boldsymbol{\gamma} \in \operatorname{int}(\Gamma_{\max})$*. Moreover, for* $p = p_1 + \ldots + p_r$, $p_i \in \mathbb{N} \cup \{0\}$ $\forall i = 1, \ldots, r$ *one has that*

$$\frac{\partial^p}{\partial \gamma_1^{p_1} \ldots \partial \gamma_r^{p_r}} \exp\{\kappa(\boldsymbol{\gamma})\} = \int_{\mathcal{Y}} t_1^{p_1}(\mathbf{y}) \ldots t_r^{p_r}(\mathbf{y}) \exp\left\{-\psi(\mathbf{y}) + t^\top(\mathbf{y})\boldsymbol{\gamma}\right\} d\tau(\mathbf{y}).$$

Using Theorem 4.6, we obtain the expected value $\overline{\mu}(\boldsymbol{\gamma})$ and the covariance matrix $\overline{\Sigma}(\boldsymbol{\gamma})$ of the sufficient statistic $t(\mathbf{Y})$ for any given $\boldsymbol{\gamma} \in \operatorname{int}(\Gamma_{\max})$:

$$\overline{\mu}(\boldsymbol{\gamma}) \equiv E_{\boldsymbol{\gamma}}[t(\mathbf{Y})] = \frac{\partial \kappa(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \begin{pmatrix} \frac{\partial \kappa(\boldsymbol{\gamma})}{\partial \gamma_1} \\ \vdots \\ \frac{\partial \kappa(\boldsymbol{\gamma})}{\partial \gamma_r} \end{pmatrix}, \tag{4.8}$$

$$\overline{\Sigma}(\boldsymbol{\gamma}) \equiv \mathrm{Var}_{\boldsymbol{\gamma}}\left[t\left(\mathbf{Y}\right)\right] = \frac{\partial^2 \kappa\left(\boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} = \begin{pmatrix} \frac{\partial^2 \kappa(\boldsymbol{\gamma})}{\partial \gamma_1 \partial \gamma_1} & \cdots & \frac{\partial^2 \kappa(\boldsymbol{\gamma})}{\partial \gamma_1 \partial \gamma_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \kappa(\boldsymbol{\gamma})}{\partial \gamma_r \partial \gamma_1} & \cdots & \frac{\partial^2 \kappa(\boldsymbol{\gamma})}{\partial \gamma_r \partial \gamma_r} \end{pmatrix} = \frac{\partial \overline{\mu}\left(\boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma}^\top} \tag{4.9}$$

(see e.g. Brown, 1986, Corollary 2.3 or Pázman, 1993, Chap. 9.1).

It can be proved that $\overline{\Sigma}(\boldsymbol{\gamma})$ is positive definite (and nonsingular) if $t_1\left(\mathbf{y}\right), \ldots, t_r\left(\mathbf{y}\right), 1$ are linearly independent functions, see e.g. Chap. 9.1 in Pázman (1993). (We say that the functions $\varphi_1\left(z\right), \ldots, \varphi_n\left(z\right)$ are linearly dependent if there are such numbers $a_1, \ldots, a_n$, not all zeros, that $\sum_{i=1}^{n} \varphi_i\left(z\right) a_i = 0 \ \forall z$.) Any pdf (or pmf) from (4.7) can be equivalently rewritten into such form that $t_1\left(\mathbf{y}\right), \ldots, t_r\left(\mathbf{y}\right), 1$ and, simultaneously, $\gamma_1\left(\boldsymbol{\vartheta}\right), \ldots, \gamma_r\left(\boldsymbol{\vartheta}\right), 1$ are linearly independent (Brown, 1986, Theorem 1.9). Such a representation will be called *minimal*, see Corollary 8.1 in Barndorff-Nielsen (1978) for this term. Throughout this thesis we will always consider the minimal representation of the exponential family.

It is easy to see that $\overline{\mu}\left[\gamma(\boldsymbol{\vartheta})\right]$ and $\overline{\Sigma}\left[\gamma(\boldsymbol{\vartheta})\right]$ are expected value and covariance matrix of the sufficient statistic for given $\boldsymbol{\vartheta}$.

Fisher information matrix in exponential families

For a more comprehensive explanation of the term Fisher information matrix we recommend Lehmann and Casella (1998).

Definition 4.7. Let $f\left(\cdot, \boldsymbol{\gamma}\right)$ be the pdf (pmf) from (4.5), then

$$M_{\boldsymbol{\gamma}} = E_{\boldsymbol{\gamma}}\left[\frac{\partial \ln f\left(\mathbf{y}, \boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma}} \frac{\partial \ln f\left(\mathbf{y}, \boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma}^\top}\right]$$

is the Fisher information matrix for the canonical parameter $\boldsymbol{\gamma}$.

Definition 4.8. Let $f\left[\cdot, \gamma(\boldsymbol{\vartheta})\right]$ be the pdf (pmf) from (4.7), then

$$M_{\boldsymbol{\vartheta}} = E_{\boldsymbol{\vartheta}}\left\{\frac{\partial \ln f\left[\mathbf{y}, \gamma(\boldsymbol{\vartheta})\right]}{\partial \boldsymbol{\vartheta}} \frac{\partial \ln f\left[\mathbf{y}, \gamma(\boldsymbol{\vartheta})\right]}{\partial \boldsymbol{\vartheta}^\top}\right\}$$

is the Fisher information matrix for the parameter $\boldsymbol{\vartheta}$.

In the minimal representation of the exponential family and assuming that $\boldsymbol{\gamma}, \gamma(\boldsymbol{\vartheta}) \in \mathrm{int}\left(\Gamma_{\max}\right)$ and that the derivatives $\frac{\partial \gamma(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^\top}$ and $\frac{\partial \overline{\mu}[\gamma(\boldsymbol{\vartheta})]}{\partial \boldsymbol{\vartheta}^\top}$ exist, the direct calculations lead to (see e.g. Pázman (1993), Eqs. 9.2.8–9.2.9)

$$M_{\boldsymbol{\gamma}} = E_{\boldsymbol{\gamma}}\left[-\frac{\partial^2 \ln f\left(\mathbf{y}, \boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top}\right] = \mathrm{Var}_{\boldsymbol{\gamma}}\left[\frac{\partial \ln f\left(\mathbf{y}, \boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma}}\right] = \overline{\Sigma}\left(\boldsymbol{\gamma}\right) = \frac{\partial^2 \kappa\left(\boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top},$$

$$
\begin{aligned}
M_{\boldsymbol{\vartheta}} &= E_{\boldsymbol{\vartheta}}\left\{-\frac{\partial^2 \ln f\left[\mathbf{y},\gamma(\boldsymbol{\vartheta})\right]}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^{\top}}\right\} \\
&= \frac{\partial\gamma^{\top}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\frac{\partial\overline{\mu}\left[\gamma(\boldsymbol{\vartheta})\right]}{\boldsymbol{\vartheta}^{\top}} \\
&= \frac{\partial\gamma^{\top}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} M_{\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\gamma(\boldsymbol{\vartheta})}\frac{\partial\gamma(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\top}} = \frac{\partial\gamma^{\top}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\overline{\Sigma}\left[\gamma(\boldsymbol{\vartheta})\right]\frac{\partial\gamma(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\top}} \\
&= \frac{\partial\overline{\mu}^{\top}\left[\gamma(\boldsymbol{\vartheta})\right]}{\partial\boldsymbol{\vartheta}}\left\{\overline{\Sigma}\left[\gamma(\boldsymbol{\vartheta})\right]\right\}^{-1}\frac{\partial\overline{\mu}\left[\gamma(\boldsymbol{\vartheta})\right]}{\partial\boldsymbol{\vartheta}^{\top}}.
\end{aligned}
\tag{4.10}
$$

It is obvious that $M_{\boldsymbol{\vartheta}} \neq M_{\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\gamma(\boldsymbol{\vartheta})}$ and, unlike the expected value $\overline{\mu}(\cdot)$ or covariance matrix $\overline{\Sigma}(\cdot)$, the Fisher information matrix depends on the parametrization.

I-divergence in exponential families

The I-divergence (Information divergence or Kullback-Leibler divergence, see Kullback and Leibler, 1951; Kullback, 1997) is often applied to measure the distance between two distributions.

Definition 4.9. The I-divergence between two pdfs (or pmfs) $f\left(\cdot,\boldsymbol{\gamma}^0\right)$ and $f(\cdot,\boldsymbol{\gamma})$ from the exponential family (4.5) is defined as

$$
I\left(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}\right) = E_{\boldsymbol{\gamma}^0}\left[\ln\frac{f\left(\mathbf{y},\boldsymbol{\gamma}^0\right)}{f\left(\mathbf{y},\boldsymbol{\gamma}\right)}\right] = \int_{\mathcal{Y}} \ln\left[\frac{f\left(\mathbf{y},\boldsymbol{\gamma}^0\right)}{f\left(\mathbf{y},\boldsymbol{\gamma}\right)}\right] f\left(\mathbf{y},\boldsymbol{\gamma}^0\right) d\tau\left(\mathbf{y}\right).
$$

As follows from (4.5), provided $\boldsymbol{\gamma}^0 \in \operatorname{int}(\Gamma_{\max})$, one can write in the exponential family that (see e.g. Pázman, 1993, Eq. 9.2.3)

$$
I\left(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}\right) = \overline{\mu}^{\top}\left(\boldsymbol{\gamma}^0\right)\left(\boldsymbol{\gamma}^0-\boldsymbol{\gamma}\right)+\kappa(\boldsymbol{\gamma})-\kappa\left(\boldsymbol{\gamma}^0\right). \tag{4.11}
$$

Notice that $I\left[\gamma\left(\boldsymbol{\vartheta}^0\right),\gamma(\boldsymbol{\vartheta})\right]$, as a function of $\boldsymbol{\vartheta}^0$ and $\boldsymbol{\vartheta}$, is the I-divergence between $f\left[\cdot,\gamma\left(\boldsymbol{\vartheta}^0\right)\right]$ and $f\left[\cdot,\gamma(\boldsymbol{\vartheta})\right]$. One can write for $\gamma(\boldsymbol{\vartheta}^0) \in \operatorname{int}(\Gamma_{\max})$ (see e.g. Pázman, 1993, Eq. 9.2.4)

$$
I\left[\gamma\left(\boldsymbol{\vartheta}^0\right),\gamma(\boldsymbol{\vartheta})\right] = \overline{\mu}^{\top}\left[\gamma\left(\boldsymbol{\vartheta}^0\right)\right]\left[\gamma\left(\boldsymbol{\vartheta}^0\right)-\gamma(\boldsymbol{\vartheta})\right]+\kappa\left[\gamma(\boldsymbol{\vartheta})\right]-\kappa\left[\gamma\left(\boldsymbol{\vartheta}^0\right)\right].
$$

The I-divergence $I(\boldsymbol{\gamma}^0,\boldsymbol{\gamma})$ is nonnegative and equals zero if and only if $f(\mathbf{y},\boldsymbol{\gamma}^0) = f(\mathbf{y},\boldsymbol{\gamma})$, see Lemma 3.1 of Kullback and Leibler (1951).

4.1.3 Generalized regression models based on exponential families

The classical linear and nonlinear models (1.4) and (1.5) considered in previous chapters do not include the case of a discrete observed variable $y(\mathbf{x})$ or the case when also

the variance component of $y(\mathbf{x})$ depends on the design point $\mathbf{x}$. Here are *generalized regression models* applicable.

Very popular are *generalized linear models*, which are well elaborated in the literature, e.g. Nelder and Wedderburn (1972); McCullagh and Nelder (1989); Dobson (1990). Although there are papers not restricted to linearity, e.g. Atkinson et al. (2014), there is a lack of complex exposition of generalized regression models in the literature. So in the thesis we rely on the above-mentioned literature, on the known properties of distributions in the exponential family, which were presented in the previous section, and on the lectures given by Pázman (Nonlinear statistical models, 2015).

In this chapter we consider the generalized regression models based on exponential families of distributions. We suppose that the set of all possible design points $\mathcal{X}$ is finite.

As before, let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, $\mathbf{x}_i \in \mathcal{X}\ \forall i = 1, \ldots, N$ denote an exact experimental design. We observe N independent random vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$ associated with the design X with the pdfs (or pmfs) of the form

$$f(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}) = \exp\left\{-\psi(\mathbf{y}_i) + t^\top(\mathbf{y}_i)\, g(\mathbf{x}_i, \boldsymbol{\theta}) - \kappa[g(\mathbf{x}_i, \boldsymbol{\theta})]\right\}, \tag{4.12}$$

which is obviously an exponential family with the canonical parameter $\boldsymbol{\gamma}_i = g(\mathbf{x}_i, \boldsymbol{\theta}) \in \Gamma_{\max}$, for any $i = 1, \ldots, N$. Here, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^m$ is the unknown model parameter and $g(\mathbf{x}_i, \boldsymbol{\theta}) = (g_1(\mathbf{x}_i, \boldsymbol{\theta}), \ldots, g_r(\mathbf{x}_i, \boldsymbol{\theta}))^\top$ is the known regression function. Unless otherwise stated, Θ is supposed to be compact set. We will assume that the function $g : \mathcal{X} \times \Theta \mapsto \mathbb{R}^r$ is three-times continuously differentiable on $\operatorname{int}(\Theta)\ \forall \mathbf{x} \in \mathcal{X}$ and that $m < rN$. Throughout this chapter we assume that (4.12) is expressed in its minimal representation and that the set $\Gamma_{\max}$ from (4.6) is open to ensure the existence of moments, see Theorem 4.6. This means that we consider exclusively the *regular* exponential families in the terminology of Barndorff-Nielsen (1978) and Brown (1986).

Not only the canonical parameter $\boldsymbol{\gamma}$, but also the usual parameter in given family or the expected value of the sufficient statistic can be parametrized by $\boldsymbol{\theta}$. These two alternative parametrizations can be applied in the praxis thanks to their straightforward interpretation. In the exponential families are the relations between the canonical parameter $\boldsymbol{\gamma}$, the usual parameter in given family, and the expected value of sufficient statistic $\overline{\mu}(\boldsymbol{\gamma})$ known (see Appendix B). This allows us to use only the canonical parametrization in our theoretical considerations.

One obtains from (4.12) that the joint distribution of the mutually independent random vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$

$$\prod_{i=1}^{N} f(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}) = \exp\left\{-\sum_{i=1}^{N}\psi(\mathbf{y}_i) + \sum_{i=1}^{N} t^\top(\mathbf{y}_i)\, g(\mathbf{x}_i, \boldsymbol{\theta}) - \sum_{i=1}^{N}\kappa[g(\mathbf{x}_i, \boldsymbol{\theta})]\right\}, \tag{4.13}$$

is obviously an exponential family.

The estimate $\hat{\boldsymbol{\theta}}_N$ for the unknown parameter $\boldsymbol{\theta}$ is then computed via maximum likelihood method:

$$\hat{\boldsymbol{\theta}}_N = \arg\max_{\boldsymbol{\theta}\in\Theta} \prod_{i=1}^{N} f(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}). \tag{4.14}$$

In exponential families of distributions and in the generalized regression models based on them, the estimate can be calculated iteratively e.g. via *the Fisher scoring method*. Numerical methods of calculations of $\hat{\boldsymbol{\theta}}_N$ are beyond the scope of this thesis and are elaborated e.g. in Nelder and Wedderburn (1972); Nelder (1975); Jørgensen (1984); Green (1984); McCullagh and Nelder (1989); Dobson (1990).

Moments, Fisher information matrix, and I-divergence in generalized regression models

When considering a generalized regression model based on exponential families, we set the regression function $g(\mathbf{x}, \boldsymbol{\theta})$ instead of the canonical parameter $\boldsymbol{\gamma}$ in (4.5) to obtain the pdf (or pmf) of the measurement $\mathbf{y}$ observed at $\mathbf{x} \in \mathcal{X}$:

$$f(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) = \exp\left\{-\psi(\mathbf{y}) + t^{\top}(\mathbf{y})\, g(\mathbf{x}, \boldsymbol{\theta}) - \kappa[g(\mathbf{x}, \boldsymbol{\theta})]\right\}, \tag{4.15}$$

which is assumed to be expressed in its minimal representation. Similarly as in linear and nonlinear regression, throughout this chapter we always suppose that any two observations $\mathbf{y}, \mathbf{y}'$ from different trials are independent. The moments for given $\mathbf{x}$ and $\boldsymbol{\theta}$ exist, since $\Gamma_{\max}$ is open, see Theorem 4.6, and we can use all results derived for (4.5), in particular, following from Eqs. (4.8)–(4.11), we can define

- the expected value of the sufficient statistic $t(\mathbf{y})$ for given $\mathbf{x}$ and $\boldsymbol{\theta}$:

$$\mu(\mathbf{x}, \boldsymbol{\theta}) = \overline{\mu}[g(\mathbf{x}, \boldsymbol{\theta})] = \left[\frac{\partial \kappa(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right]_{\boldsymbol{\gamma}=g(\mathbf{x},\boldsymbol{\theta})}, \tag{4.16}$$

- the covariance matrix of the sufficient statistic $t(\mathbf{y})$ for given $\mathbf{x}$ and $\boldsymbol{\theta}$:

$$\Sigma(\mathbf{x}, \boldsymbol{\theta}) = \overline{\Sigma}[g(\mathbf{x}, \boldsymbol{\theta})] = \left[\frac{\partial^2 \kappa(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^{\top}}\right]_{\boldsymbol{\gamma}=g(\mathbf{x},\boldsymbol{\theta})},$$

- the elementary I-divergence for given $\mathbf{x}$:

$$\begin{aligned} I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right) &= E_{\boldsymbol{\theta}^0}\left[\ln \frac{f\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}^0\right)}{f(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})}\right] = I\left[g\left(\mathbf{x}, \boldsymbol{\theta}^0\right), g(\mathbf{x}, \boldsymbol{\theta})\right] \\ &= \mu^{\top}\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\left[g\left(\mathbf{x}, \boldsymbol{\theta}^0\right) - g(\mathbf{x}, \boldsymbol{\theta})\right] + \kappa[g(\mathbf{x}, \boldsymbol{\theta})] - \kappa\left[g\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\right], \end{aligned} \tag{4.17}$$

- the I-divergence for the exact design $X=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$:

$$I_X\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)=E_{\boldsymbol{\theta}^0}\left[\ln\frac{\prod_{i=1}^N f\left(\mathbf{y}_i,\mathbf{x}_i,\boldsymbol{\theta}^0\right)}{\prod_{i=1}^N f\left(\mathbf{y}_i,\mathbf{x}_i,\boldsymbol{\theta}\right)}\right]=\sum_{i=1}^N E_{\boldsymbol{\theta}^0}\left[\ln\frac{f\left(\mathbf{y}_i,\mathbf{x}_i,\boldsymbol{\theta}^0\right)}{f\left(\mathbf{y}_i,\mathbf{x}_i,\boldsymbol{\theta}\right)}\right]$$
$$=\sum_{i=1}^N I_{\mathbf{x}_i}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right),$$

- the elementary Fisher information matrix for the parameter $\boldsymbol{\theta}$ is defined as $M\left(\mathbf{x},\boldsymbol{\theta}\right)\equiv E_{\boldsymbol{\theta}}\left[\frac{\partial\ln f(\mathbf{y},\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\ln f(\mathbf{y},\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right]$. One has the following relations:

$$\begin{aligned}M\left(\mathbf{x},\boldsymbol{\theta}\right)&=E_{\boldsymbol{\theta}}\left[-\frac{\partial^2\ln f\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right]\\&=\frac{\partial g^\top\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\frac{\partial\mu\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^\top}\\&=\frac{\partial g^\top\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}M_{\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=g(\mathbf{x},\boldsymbol{\theta})}\frac{\partial g\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^\top}=\frac{\partial g^\top\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\Sigma\left(\mathbf{x},\boldsymbol{\theta}\right)\frac{\partial g\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^\top}\\&=\frac{\partial\mu^\top\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\Sigma^{-1}\left(\mathbf{x},\boldsymbol{\theta}\right)\frac{\partial\mu\left(\mathbf{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^\top},\end{aligned}\tag{4.18}$$

  provided that $\mu(\mathbf{x},\boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}\in\operatorname{int}(\Theta)$,

- the Fisher information matrix associated with the exact design $X=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$: $M_X\left(\boldsymbol{\theta}\right)=E_{\boldsymbol{\theta}}\left[-\frac{\partial^2\ln\prod_{i=1}^N f(\mathbf{y}_i,\mathbf{x}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right]=\sum_{i=1}^N E_{\boldsymbol{\theta}}\left[-\frac{\partial^2\ln f(\mathbf{y}_i,\mathbf{x}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right]=\sum_{i=1}^N M\left(\mathbf{x}_i,\boldsymbol{\theta}\right)$.

If the random variable ε in the nonlinear regression model (1.5) or in the linear regression model (1.4) is normally distributed with zero mean and unknown constant variance $\sigma^2>0$, then the models (1.5) and (1.4) will be called *normal nonlinear model* and *normal linear model*, respectively.

Remark 4.10. The nonlinear regression model (1.5) with normally distributed random errors with zero mean and unit variance (i.e. $y\left(\mathbf{x}\right)\sim\mathcal{N}\left(\eta\left(\mathbf{x},\boldsymbol{\theta}\right),1\right)$) can be interpreted as a generalized regression model. One can see in the Appendix B.8 that the usual parameter, the canonical parameter and the expected value of the sufficient statistic coincide and the corresponding regression function is given by the relation $g\left(\mathbf{x},\boldsymbol{\theta}\right)=\eta\left(\mathbf{x},\boldsymbol{\theta}\right)$. One has $I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)=1/2\left[\eta\left(\mathbf{x},\boldsymbol{\theta}^0\right)-\eta\left(\mathbf{x},\boldsymbol{\theta}\right)\right]^2$ and $M\left(\mathbf{x},\boldsymbol{\theta}\right)=\frac{\partial g^\top(\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Sigma\left(\mathbf{x},\boldsymbol{\theta}\right)\frac{\partial g(\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}=\frac{\partial\eta(\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\eta(\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}$ for the elementary I-divergence and elementary information matrix, respectively. If, moreover, the relation between the regressors and the parameters is linear, i.e. if $\eta\left(\mathbf{x},\boldsymbol{\theta}\right)=\mathbf{f}^\top\left(\mathbf{x}\right)\boldsymbol{\theta}$, we obtain a normal linear regression model as a special case of generalized regression models and $M\left(\mathbf{x}\right)=\mathbf{f}\left(\mathbf{x}\right)\mathbf{f}^\top\left(\mathbf{x}\right)$, $I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)=1/2\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}\right)^\top M\left(\mathbf{x}\right)\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}\right)$.

Asymptotic properties of maximum likelihood estimate in generalized regression models

Consider an exact experimental design $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, where $\mathbf{x}_1, \ldots, \mathbf{x}_N$ is a random sample from the distribution given by the approximative design ξ. It can be shown (see Sects. 4.2–4.3 of Pronzato and Pázman, 2013) that, under some assumptions, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_N$ from (4.14) is asymptotically (for large N) normally distributed with mean equal to the true parameter value $\bar{\boldsymbol{\theta}}$ and with covariance matrix $\frac{1}{N}\left[M(\xi, \bar{\boldsymbol{\theta}})\right]^{-1}$, where $M(\xi, \bar{\boldsymbol{\theta}})$ is assumed to be nonsingular information matrix associated with design ξ defined as

$$M(\xi, \boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathcal{X}} M(\mathbf{x}, \boldsymbol{\theta}) \xi(\mathbf{x})$$

with $M(\mathbf{x}, \boldsymbol{\theta})$ from (4.18).

4.2 Experimental design in generalized regression models

The design issues in generalized regression models are studied e.g. in Atkinson et al. (2014) and in generalized linear models e.g. in papers Atkinson and Woods (2015) or Khuri et al. (2006).

In this chapter we consider two different approaches to optimal experimental design in generalized regression models based on exponential families:

1. The approach which uses the asymptotic properties of maximum likelihood estimate as in Atkinson et al. (2014) and
2. the approach similar to Pázman and Pronzato (2014), who introduced the extended optimality criteria which lead simultaneously to stable and precise estimates in nonlinear regression models. Our extension of these results is described in the next sections of this chapter.

We remind that throughout Chap. 4 we always assume that the design space $\mathcal{X}$ is finite.

4.2.1 Designing experiments in generalized regression models using the asymptotic properties of maximum likelihood estimate

The main idea of this approach is to maximize the proper function of Fisher information matrix $M(\xi, \boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathcal{X}} M(\mathbf{x}, \boldsymbol{\theta}) \xi(\mathbf{x})$ as in Atkinson et al. (2014). The asymptotic

properties of maximum likelihood estimate indicated at the end of Sect. 4.1.3 are then applied.

Unlike the classical nonlinear regression model (1.5), where the rank of elementary information matrix equals one, in generalized regression models we have

$$M(\mathbf{x},\boldsymbol{\theta}) = F(\mathbf{x},\boldsymbol{\theta})\, F^{\top}(\mathbf{x},\boldsymbol{\theta}),$$

where $F(\mathbf{x},\boldsymbol{\theta}) = \frac{\partial g^{\top}(\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Sigma^{1/2}(\mathbf{x},\boldsymbol{\theta})$, see (4.18), is $m \times r$ matrix and its rank is not necessarily equal to one (it equals one if the sufficient statistic $t(\mathbf{y})$ is one-dimensional).

In generalized regression models we can also apply the results of Chap. 2 and compute locally optimal designs via linear programming after the proper reformulation of the optimality criterion into the form

$$\forall\, \xi \in \Xi^* \quad \Phi_{\mathrm{loc}}(\xi) = \phi\left(\xi, \boldsymbol{\theta}^0\right) = \min_{\zeta \in \Xi^*} \sum_{\mathbf{x} \in \mathcal{X}} H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi(x), \tag{4.19}$$

where the function $H(\cdot,\cdot,\cdot)$ and the set Ξ^* depend on the criterion which is reformulated:

D-optimality criterion $\Xi^* = \{\zeta : M(\zeta,\boldsymbol{\theta}) \text{ is nonsingular}\}$

$$H(\zeta,\mathbf{x},\boldsymbol{\theta}) = \frac{\det^{1/m}[M(\zeta,\boldsymbol{\theta})]}{m} \mathrm{tr}\Big[F^{\top}(\mathbf{x},\boldsymbol{\theta})\, M^{-1}(\zeta,\boldsymbol{\theta})\, F(\mathbf{x},\boldsymbol{\theta})\Big],$$

A-optimality criterion $\Xi^* = \{\zeta : M(\zeta,\boldsymbol{\theta}) \text{ is nonsingular}\}$

$$H(\zeta,\mathbf{x},\boldsymbol{\theta}) = \frac{\mathrm{tr}\Big[F^{\top}(\mathbf{x},\boldsymbol{\theta})\, M^{-2}(\zeta,\boldsymbol{\theta})\, F(\mathbf{x},\boldsymbol{\theta})\Big]}{\mathrm{tr}^2[M^{-1}(\zeta,\boldsymbol{\theta})]},$$

E_k-optimality criterion $\Xi^* = \Xi$

$$H(\zeta,\mathbf{x},\boldsymbol{\theta}) = \mathrm{tr}\Big[F^{\top}(\mathbf{x},\boldsymbol{\theta})\, P^{(k)}(\zeta,\boldsymbol{\theta})\, F(\mathbf{x},\boldsymbol{\theta})\Big],$$

where $P^{(k)}(\zeta,\boldsymbol{\theta}) = \sum_{i=1}^{k} \mathbf{u}_i[M(\zeta,\boldsymbol{\theta})]\, \mathbf{u}_i^{\top}[M(\zeta,\boldsymbol{\theta})]$ is a k-dimensional orthogonal projector and $\mathbf{u}_i[M(\zeta,\boldsymbol{\theta})]$ $i = 1,\ldots,k$ are the orthonormal eigenvectors of matrix $M(\zeta,\boldsymbol{\theta})$ corresponding to its k smallest eigenvalues.

The reformulation (4.19) can be justified in the same way as in Chap. 2. The maximization of criterion Φ_{loc} then corresponds to an LP problem with infinitely many linear constraints, which leads to the similar algorithm as introduced in Chap. 2.

To avoid the undesirable dependence of the information matrix on the unknown parameter $\boldsymbol{\theta}$, one may also apply the maximin optimality criterion (1.9), AVE optimality criterion (1.10), or criterion based on the conditional value at risk (CVaR) as a compromise (see Chap. 3).

4.3 The problem of stability and identifiability of maximum likelihood estimate. The extended optimality criteria.

For a detailed discussion on estimability and identifiability of parameters in classical nonlinear regression we refer to Chap. 7 of Pronzato and Pázman (2013).

It turns out to be very important to describe and identify the possible instabilities which may appear in maximum likelihood estimation in generalized regression models. Especially for a normal nonlinear regression we observe the same instabilities as described in Pázman and Pronzato (2014) or in Chap. 7 of Pronzato and Pázman (2013), since in the case of normally distributed random errors, the maximum likelihood method and the method of least squares lead to the same estimates.

Let $\bar{\boldsymbol{\theta}}$ be the true and unknown parameter value and let $\hat{\boldsymbol{\theta}}_N$ be its maximum likelihood estimate (4.14) based on N independent measurements in $\mathbf{x}_1, \ldots, \mathbf{x}_N$ such that the relative frequency of $\mathbf{x}$ within $\mathbf{x}_1, \ldots, \mathbf{x}_N$ tends for $N \to \infty$ to $\xi(\mathbf{x})$. Suppose that the nominal parameter value $\boldsymbol{\theta}^0$ is allocated in the neighbourhood of $\bar{\boldsymbol{\theta}}$. Then the variability of the estimator $\hat{\boldsymbol{\theta}}_N$ near $\bar{\boldsymbol{\theta}}$ is well expressed via the information matrix $M\left(\xi, \boldsymbol{\theta}^0\right)$, since $M^{-1}\left(\xi, \bar{\boldsymbol{\theta}}\right)$ is proportional to the asymptotic covariance matrix of the estimate $\hat{\boldsymbol{\theta}}_N$ (see Sect. 4.1.3). The experimental design which maximizes the classical local optimality criterion (1.8) ensures that the parameter $\boldsymbol{\theta}$ is locally well identified in the neighbourhood of $\boldsymbol{\theta}^0$ (i.e. the covariance matrix of $\hat{\boldsymbol{\theta}}_N$ is "small").

We are interested in the global identifiability (or stability) of the parameter that is related to points $\boldsymbol{\theta}$ distant from $\bar{\boldsymbol{\theta}}$. The problem appears when for such $\boldsymbol{\theta}$ the likelihood function $L(\boldsymbol{\theta}) = \prod_{i=1}^N f(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta})$ is very close to $L\left(\bar{\boldsymbol{\theta}}\right) = \prod_{i=1}^N f\left(\mathbf{y}_i, \mathbf{x}_i, \bar{\boldsymbol{\theta}}\right)$, i.e. when the difference $\left|\ln\left[L\left(\bar{\boldsymbol{\theta}}\right)\right] - \ln\left[L(\boldsymbol{\theta})\right]\right|$ is very small. In such cases, the maximum of the likelihood function can be attained at $\boldsymbol{\theta}$, a point distant from the true parameter value $\bar{\boldsymbol{\theta}}$. It follows from (4.13) and (4.16) that this instability may appear when the "canonical surface"

$$\mathscr{K} = \left\{\left(g^\top(\mathbf{x}_1, \boldsymbol{\theta}), \ldots, g^\top(\mathbf{x}_N, \boldsymbol{\theta})\right)^\top, \boldsymbol{\theta} \in \Theta\right\}$$

or "expectation surface"

$$\mathscr{E} = \left\{\left(\mu^\top(\mathbf{x}_1, \boldsymbol{\theta}), \ldots, \mu^\top(\mathbf{x}_N, \boldsymbol{\theta})\right)^\top, \boldsymbol{\theta} \in \Theta\right\}$$

are "nearly overlapping", e.g. when they are shaped as in Figure 4.1 on page 97. The importance of the canonical and expectation surface in curved exponential families was emphasized by Efron (1978).

The problem of both global and local identifiability can be well described via the I-divergence. Directly from (4.17) we obtain the following remark.

Remark 4.11. Consider an exact experimental design $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and an approximative design ξ. In the generalized regression model with the density of measurements (4.15) one has

$$I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right) = E_{\boldsymbol{\theta}^0}\left[\ln f\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}^0\right) - \ln f\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}\right)\right] \ \forall\, \mathbf{x} \in \mathcal{X},$$

$$I_X\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right) = E_{\boldsymbol{\theta}^0}\left[\ln \prod_{i=1}^{N} f\left(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}^0\right) - \ln \prod_{i=1}^{N} f\left(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}\right)\right],$$

and, obviously,

$$\sum_{\mathbf{x} \in \mathcal{X}} I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right) \xi(\mathbf{x}) = E_{\boldsymbol{\theta}^0}\left\{\sum_{\mathbf{x} \in \mathcal{X}} \left[\ln f\left(\mathbf{y}(\mathbf{x}), \mathbf{x}, \boldsymbol{\theta}^0\right) - \ln f\left(\mathbf{y}(\mathbf{x}), \mathbf{x}, \boldsymbol{\theta}\right)\right] \xi(\mathbf{x})\right\}.$$

At the design stage of the experiment, neither the observed values $\mathbf{y}_1, \ldots, \mathbf{y}_N$ nor the difference $\ln\left[\prod_{i=1}^N f\left(\mathbf{y}_i, \mathbf{x}_i, \bar{\boldsymbol{\theta}}\right)\right] - \ln\left[\prod_{i=1}^N f\left(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}\right)\right]$ (which we prefer to be as large as possible for $\boldsymbol{\theta}$ distant from $\bar{\boldsymbol{\theta}}$) are known. According to Remark 4.11 and under the assumption that $\boldsymbol{\theta}^0$ is in the neighbourhood of $\bar{\boldsymbol{\theta}}$, one can use the I-divergence $\sum_{\mathbf{x} \in \mathcal{X}} I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right) \xi(\mathbf{x})$ to express the global identifiability of the parameter at $\boldsymbol{\theta}^0$ associated with the design ξ.

On the other hand, the I-divergence reflects also the variability of the estimate $\hat{\boldsymbol{\theta}}_N$ near $\boldsymbol{\theta}^0$ as a consequence of the following lemma.

Lemma 4.12. *Let the third order derivatives* $\frac{\partial^3 I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right)}{\partial \theta_h \partial \theta_i \partial \theta_j}$ *be bounded for any* $\mathbf{x} \in \mathcal{X}$ *and any* $\boldsymbol{\theta} \in \operatorname{int}(\Theta)$ $\forall h, i, j \in \{1, \ldots, m\}$. *Then, in the generalized regression model with the density of measurements (4.15),*

$$I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right) = \frac{1}{2}\left(\boldsymbol{\theta}^0 - \boldsymbol{\theta}\right)^{\top} M\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\left(\boldsymbol{\theta}^0 - \boldsymbol{\theta}\right) + O\left(\left\|\boldsymbol{\theta}^0 - \boldsymbol{\theta}\right\|_{\ell(2)}^3\right).$$

The big O notation $O(\cdot)$ *here describes the behaviour of a given function when* $\left\|\boldsymbol{\theta}^0 - \boldsymbol{\theta}\right\|_{\ell(2)} \to 0$.

Remark 4.13. One writes that $\varphi(z) = O[\varrho(z)]$ for $z \to 0$ for some functions $\varrho(z)$ and $\varphi(z)$ if and only if $\exists \Delta > 0, Q > 0$ such that $|\varphi(z)| \leq Q\,|\varrho(z)|$ for $|z| < \Delta$. This also implies that if for some $n \in \mathbb{N} \cup \{0\}$ $\varphi_n(z)$ is $O(z^n)$ and $\varphi_{n+1}(z)$ is $O\left(z^{n+1}\right)$ for $z \to 0$, then $\varphi_n(z) + \varphi_{n+1}(z) = O(z^n)$, $\varphi_n(z)\,\varphi_{n+1}(z) = O\left(z^{2n+1}\right)$ for $z \to 0$.

Proof of Lemma 4.12. Equations (4.16)–(4.18) imply that

$$I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}^0\right) = 0, \quad \left[\frac{\partial I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} = \mathbf{0}, \quad \left[\frac{\partial^2 I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}}\right]_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} = M\left(\mathbf{x}, \boldsymbol{\theta}^0\right).$$

Then by using the Taylor expansion of $I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)$ at $\boldsymbol{\theta}^0$ (see e.g. Trench, 2003, Theorem 5.4.8) can the lemma be proved.

Notice that in the Taylor expansion we used the Lagrange reminder (Abramowitz and Stegun, 1972, p. 880):

$$\frac{1}{3!}\sum_{h,i,j}\left(\theta_h-\theta_h^0\right)\left(\theta_i-\theta_i^0\right)\left(\theta_j-\theta_j^0\right)\left[\frac{\partial^3 I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)}{\partial\theta_h\partial\theta_i\partial\theta_j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star},$$

for some $\boldsymbol{\theta}^\star$ on the line segment connecting $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}$. The boundedness of the derivatives ensures the existence of such Q_0 that

$$\left|\left[\frac{\partial^3 I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)}{\partial\theta_h\partial\theta_i\partial\theta_j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\right|\leq Q_0$$

for any $h,i,j\in\{1,\ldots,m\}$. It follows that there is $Q>0$ such that

$$\left|\frac{1}{3!}\sum_{h,i,j}\left(\theta_h-\theta_h^0\right)\left(\theta_i-\theta_i^0\right)\left(\theta_j-\theta_j^0\right)\left[\frac{\partial^3 I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)}{\partial\theta_h\partial\theta_i\partial\theta_j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\right|\leq$$
$$\leq\frac{Q_0}{3!}\sum_{h,i,j}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{\ell(2)}^3\leq Q\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{\ell(2)}^3,$$

which justifies that the last addend in the Taylor expansion is $O\left(\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{\ell(2)}^3\right)$. □

Lemma 4.12 implies for the information matrix $M\left(\xi,\boldsymbol{\theta}^0\right)=\sum_{\mathbf{x}\in\mathcal{X}}M\left(\mathbf{x},\boldsymbol{\theta}^0\right)\xi\left(\mathbf{x}\right)$ that

$$\sum_{\mathbf{x}\in\mathcal{X}}I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)=\frac{1}{2}\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}\right)^\top M\left(\xi,\boldsymbol{\theta}^0\right)\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}\right)+O\left(\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}\right\|_{\ell(2)}^3\right).\tag{4.20}$$

4.3.1 The general definition of extended optimality criteria

Let us introduce a very general definition of extended optimality criterion which is intended to be maximized with respect to $\xi\in\Xi$.

Definition 4.14. Let $k\in\mathbb{N}$ be a given number and $K\geq 0$ be a tuning constant. The extended optimality criterion for a given nominal parameter value $\boldsymbol{\theta}^0$ is defined as:

$$\phi_\rho^{ext}\left(\xi,\boldsymbol{\theta}^0\right)\equiv\inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right],\tag{4.21}$$

where $\rho:\mathbb{R}^{m\times(k+1)}\to\mathbb{R}$, $\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\mapsto\rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)$ is a distance between a k-tuple of points $\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k$ and the nominal value $\boldsymbol{\theta}^0$ in the parametric space Θ.

The choice of the distance ρ and of the number $k \in \mathbb{N}$ directly determines the extension of which criterion is dealt with, as we will see in Chaps. 4.4 and 4.5.

The term *extended optimality criteria* is from Pázman and Pronzato (2014), who introduced the extended criteria of E-, **c**-, and G-optimality in classical the nonlinear regression models (1.5). They considered the case $k = 1$ and instead of the double I-divergence in the numerator, they put $\left[\eta\left(\mathbf{x}, \boldsymbol{\theta}^0\right) - \eta\left(\mathbf{x}, \boldsymbol{\theta}\right)\right]^2$. They also justified the relation of the extended criteria to the classical criteria of E-, **c**-, and G-optimality.

In the thesis we always denote the extended optimality criteria by *ext* in the superscript to avoid possible confusions with the classical (not extended) optimality criteria.

The extended criteria (4.21) are based on properly "standardized" I-divergences. Notice that the I-divergence has already been used in the optimal experimental design by López-Fidalgo et al. (2007), who introduced the criterion of KL-optimality for (different) purposes of model discrimination.

For $K = 0$ (see Sect. 4.6.1 for more properties of the tuning parameter) is the interpretation of the criterion in (4.21) as follows: the aim of the experimental design is to find $\xi^\star$ which maximizes the minimal ratio of the summary I-divergence $\sum_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)$ to the distance $\rho\left(\boldsymbol{\theta}^0; \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k\right)$. The ratio is minimized over the set of all possible ordered k-tuples of vectors $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k$, hence, the optimal design $\xi^\star$ prevent such situation when the summary I-divergence is small for $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k$ very distant from $\boldsymbol{\theta}^0$, and, subsequently, the probability of false maximum likelihood estimate is minimized (see Remark 4.11). Pázman and Pronzato (2014) considered only the case $k = 1$ when one searches for one "bad" $\boldsymbol{\theta}$ minimizing the above mentioned ratio. If $k > 1$, we can consider more than one "bad" $\boldsymbol{\theta}$, and hence the resulting criterion is more "robust" in this sense.

Nevertheless, the interpretation is clearer for $k = 1$ and we also have to mention the computational complexity increasing with k.

4.4 The extension of pseudonorm optimality criteria

Consider the class of (local) **pseudonorm criteria**

$$\phi_{|||\cdot|||_P}\left(\xi, \boldsymbol{\theta}\right) = \inf_{A\in\mathbb{R}^{m\times k}:|||A|||_P=1} \operatorname{tr}\left[A^\top M\left(\xi, \boldsymbol{\theta}\right) A\right], \tag{4.22}$$

where m is equal to the dimension of the unknown parameter $\boldsymbol{\theta}$, $k \in \mathbb{N}$ is a given number and $|||\cdot|||_P$ is an arbitrary pseudonorm on $\mathbb{R}^{m\times k}$ not identically equal to zero. The criterion $\phi_{|||\cdot|||_P}$ is concave and positively homogeneous in ξ, since it is defined as infimum of functions linear in ξ.

When $|||\cdot|||_P$ corresponds to an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^{m\times k}$, then the infimum in (4.22) is reached, because the set $\left\{A\in\mathbb{R}^{m\times k}:\|A\|=1\right\}$ is compact and the mapping $A\mapsto \operatorname{tr}\left[A^\top M(\xi,\boldsymbol{\theta})A\right]$ is a continuous function in A on $\mathbb{R}^{m\times k}$. Moreover, in this case, the criterion (4.22) is equal to the information function of Dette et al. (1995), who worked with the corresponding convex class of criteria, also called minimax criteria. Here we use the term pseudonorm criteria for $\phi_{|||\cdot|||_P}$ from (4.22) to distinguish between the maximin criterion $\Phi_{\min}$ defined in (1.9).

Lemma 4.15. *Suppose that $|||\cdot|||_P$ is a norm $\|\cdot\|$ on $\mathbb{R}^{m\times k}$. When the matrix $M(\xi,\boldsymbol{\theta})$ is singular, then the criterion $\phi_{\|\cdot\|}(\xi,\boldsymbol{\theta})$ equals 0.*

Proof. When $M(\xi,\boldsymbol{\theta})$ is singular, then there is a vector $\mathbf{a}\neq\mathbf{0}\in\mathbb{R}^m$ such that $M(\xi,\boldsymbol{\theta})\mathbf{a}=\mathbf{0}$ and we set $A=(\mathbf{a},\ldots,\mathbf{a})\in\mathbb{R}^{m\times k}$, obviously $\|A\|\neq 0$ due to Def. 4.1. One sees that $0=\frac{\operatorname{tr}\left[A^\top M(\xi,\boldsymbol{\theta})A\right]}{\|A\|^2}=\inf_{A\in\mathbb{R}^{m\times k}:\|A\|=1}\operatorname{tr}\left[A^\top M(\xi,\boldsymbol{\theta})A\right]=\phi_{\|\cdot\|}$. □

If the distance ρ in (4.21) is given through a pseudonorm $|||\cdot|||_P$ so that $\rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)=\left|\left|\left|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^k\right)\right|\right|\right|_P$, we deal with **the extended pseudonorm criteria**

$$\phi^{ext}_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right)=\inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{\left|\left|\left|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^k\right)\right|\right|\right|_P^2}+K\right], \tag{4.23}$$

and their relation to the classical pseudonorm criteria (4.22) is indicated in Theorems 4.16, 4.17, and 4.18. Analogical statements about the extended criteria of E-, **c**-, and G-optimality in classical nonlinear regression were pointed out in Pázman and Pronzato (2014) and in Chap. 7 of Pronzato and Pázman (2013).

Theorem 4.16. *Let $\Theta=\mathbb{R}^m$ and $\rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)=|||\mathbf{T}_0-\mathbf{T}|||_P$, where $\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\mathbb{R}^{m\times k}$, $\mathbf{T}_0=\left(\boldsymbol{\theta}^0,\ldots,\boldsymbol{\theta}^0\right)\in\mathbb{R}^{m\times k}$, with a given nominal parameter value $\boldsymbol{\theta}^0$ and an arbitrary pseudonorm $|||\cdot|||_P$ on $\mathbb{R}^{m\times k}$ not identically equal to zero. In the normal linear regression model (1.4) with $Var[y(\mathbf{x})]=1\ \forall\mathbf{x}\in\mathcal{X}$ and with an information matrix $M(\xi)=\sum_{\mathbf{x}\in\mathcal{X}}\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})\xi(\mathbf{x})$, the following equality holds for any $K\geq 0$*

$$\inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right]=$$
$$=\inf_{A\in\mathbb{R}^{m\times k}:|||A|||_P=1}\operatorname{tr}\left[A^\top M(\xi)A\right].$$

Proof. Using Remark 4.10 one obtains

$$\begin{aligned}
&\inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\Theta^k} \sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right]\\
&=\inf_{\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\mathbb{R}^{m\times k}} \sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right)^{\top} M\left(\mathbf{x}\right)\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{|||\mathbf{T}_0-\mathbf{T}|||_P^2}+K\right]\\
&=\inf_{\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\mathbb{R}^{m\times k}} \sum_{i=1}^{k}\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right)^{\top} M\left(\xi\right)\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right)\left[\frac{1}{|||\mathbf{T}_0-\mathbf{T}|||_P^2}+K\right]\\
&=\inf_{\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\mathbb{R}^{m\times k}} \operatorname{tr}\left[\left(\mathbf{T}^0-\mathbf{T}\right)^{\top} M\left(\xi\right)\left(\mathbf{T}^0-\mathbf{T}\right)\right]\left[\frac{1}{|||\mathbf{T}_0-\mathbf{T}|||_P^2}+K\right]\\
&=\inf_{\delta>0}\inf_{A\in\mathbb{R}^{m\times k}:|||A|||_P=\delta}\operatorname{tr}\left[\frac{A^{\top}}{\delta}M\left(\xi\right)\frac{A}{\delta}\right]\left[K\delta^2+1\right]\\
&=\inf_{\delta>0}\inf_{A\in\mathbb{R}^{m\times k}:|||A|||_P=1}\operatorname{tr}\left[A^{\top}M\left(\xi\right)A\right]\left[K\delta^2+1\right]\\
&=\inf_{A\in\mathbb{R}^{m\times k}:|||A|||_P=1}\operatorname{tr}\left[A^{\top}M\left(\xi\right)A\right].
\end{aligned}$$

□

Theorem 4.17. *Suppose that the assumptions of Lemma 4.12 hold. Denote by* $\mathscr{B}\left(\boldsymbol{\theta}^0,r\right)$ *an* m*-dimensional ball centred at* $\boldsymbol{\theta}^0$ *with a diameter* $r>0$*, i.e.* $\mathscr{B}\left(\boldsymbol{\theta}^0,r\right)=\left\{\boldsymbol{\theta}:\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{\ell(2)}\leq r\right\}$*. Let* $M\left(\xi,\boldsymbol{\theta}^0\right)=\sum_{\mathbf{x}\in\mathcal{X}}M\left(\mathbf{x},\boldsymbol{\theta}^0\right)\xi\left(\mathbf{x}\right)$ *be nonsingular, and* $\rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)=|||\mathbf{T}_0-\mathbf{T}|||_P$*, where* $\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\mathbb{R}^{m\times k}$*,* $\mathbf{T}_0=\left(\boldsymbol{\theta}^0,\ldots,\boldsymbol{\theta}^0\right)\in\mathbb{R}^{m\times k}$ *and where* $|||\cdot|||_P$ *is an arbitrary pseudonorm on* $\mathbb{R}^{m\times k}$*. Moreover, suppose that there is a matrix* $A_P\in\mathbb{R}^{m\times k}$ *such that* $|||A_P|||_P=1$ *and* $\operatorname{tr}\left[A_P^{\top}M\left(\xi,\boldsymbol{\theta}^0\right)A_P\right]=\phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right)$*. Then for any* $K\geq 0$

$$\lim_{r\to 0}\inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\mathscr{B}^k(\boldsymbol{\theta}^0,r)}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right]=\phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right).$$

For the existence of the matrix A_P in the assumptions of Theorem 4.17 it suffices that Theorem A1 from Appendix can be applied to the pseudonorm criterion or that $|||\cdot|||_P$ is a norm (since then the infimum in (4.22) is always attained). In the thesis we consider only such pseudonorm criteria for which the matrix A_P does exist. Namely, the criteria of E-, MV-, **c**-, and G-optimality in Sect. 4.4.1, the criterion of A-optimality in Sect. 4.4.2 and the criterion of L-optimality in Sect. 4.4.3.

Proof. Using (4.20) one obtains for $\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)$ that

$$\begin{aligned}
0 &\leq \sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right] \\
&= \sum_{i=1}^{k}\left[\frac{\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right)^{\top} M\left(\xi,\boldsymbol{\theta}^0\right)\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right)}{\left|\left|\left|\mathbf{T}_0-\mathbf{T}\right|\right|\right|_P^2}+\frac{O\left(\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^3\right)}{\left|\left|\left|\mathbf{T}_0-\mathbf{T}\right|\right|\right|_P^2}\right]\left[1+\left|\left|\left|\mathbf{T}_0-\mathbf{T}\right|\right|\right|_P^2 K\right] \\
&= \left[\frac{\operatorname{tr}\left[\left(\mathbf{T}_0-\mathbf{T}\right)^{\top} M\left(\xi,\boldsymbol{\theta}^0\right)\left(\mathbf{T}_0-\mathbf{T}\right)\right]}{\left|\left|\left|\mathbf{T}_0-\mathbf{T}\right|\right|\right|_P^2}+\frac{O\left(\left\|\mathbf{T}_0-\mathbf{T}\right\|_F^3\right)}{\left|\left|\left|\mathbf{T}_0-\mathbf{T}\right|\right|\right|_P^2}\right]\left[1+\left|\left|\left|\mathbf{T}_0-\mathbf{T}\right|\right|\right|_P^2 K\right],
\end{aligned} \tag{4.24}$$

where (as a direct consequence of Lemma 4.12) one has that $O\left(\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^3\right)$ is big O notation for $\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}\to 0$, i.e. $\exists\Delta_i>0, Q_i>0$ such that if $\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}<\Delta_i$, then $\left|O\left(\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^3\right)\right|\leq Q_i\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^3$.

Moreover, there is $\Delta=\min\left\{\Delta_1,\ldots,\Delta_k\right\}$ and $Q=\sum_{i=1}^{k}Q_i$ such that if $\left\|\mathbf{T}_0-\mathbf{T}\right\|_F\leq\Delta$, then we have $\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}\leq\left\|\mathbf{T}_0-\mathbf{T}\right\|_F\leq\Delta\leq\Delta_i$ for any $i=1,\ldots,k$, where we used the definition of the Frobenius norm (4.2). Hence one obtains

$$\left|\sum_{i=1}^{k}O\left(\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^3\right)\right|\leq\sum_{i=1}^{k}Q_i\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^3\leq Q\left\|\mathbf{T}_0-\mathbf{T}\right\|_F^3$$

and $\sum_{i=1}^{k}O\left(\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^3\right)$ is $O\left(\left\|\mathbf{T}_0-\mathbf{T}\right\|_F^3\right)$ for $\left\|\mathbf{T}_0-\mathbf{T}\right\|_F\to 0$, which justifies the usage of $O\left(\left\|\mathbf{T}_0-\mathbf{T}\right\|_F^3\right)$ in (4.24).

Denote $\mathcal{A}_r=\left\{A:\left\|A_{\cdot i}\right\|_{\ell(2)}\leq r\ \forall i=1,\ldots,k\right\}$. Using the definition of the Frobenius norm (4.2) one sees that $\left\|A\right\|_F=\sqrt{\left\|A_{\cdot 1}\right\|_{\ell(2)}^2+\ldots+\left\|A_{\cdot k}\right\|_{\ell(2)}^2}\leq\sqrt{k}r$ for any $A\in\mathcal{A}_r$, and hence $O\left(\left\|A\right\|_F\right)=O\left(r\right)$ for $r\to 0$ and for $A\in\mathcal{A}_r$. In addition, it follows from Lemma 4.3 that for some $a>0$

$$\left|\left|\left|A\right|\right|\right|_P\leq\left\|A\right\|_F/a\leq\sqrt{k}r/a$$

for any $A\in\mathcal{A}_r$ and hence $\left|\left|\left|A\right|\right|\right|_P^2$ is $O\left(r^2\right)$ for $r\to 0$ and $A\in\mathcal{A}_r$. Denote $A_r=r\frac{A_P}{\left\|A_P\right\|_F}$.

Obviously $A_r \in \mathcal{A}_r$, $|||A_r|||_P = \frac{r}{\|A_P\|_F} > 0$ and this together with (4.24) leads to

$$
\begin{aligned}
0 &\leq \lim_{r\to 0} \inf_{\mathbf{T}=(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\mathscr{B}^k(\boldsymbol{\theta}^0,r)} \sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{|||\mathbf{T}_0-\mathbf{T}|||_P^2}+K\right] \\
&= \lim_{r\to 0}\inf_{A\in\mathcal{A}_r}\left|\frac{\mathrm{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right]+O\left(\|A\|_F^3\right)}{|||A|||_P^2}\right|\left[1+|||A|||_P^2K\right] \\
&\leq \lim_{r\to 0}\left|\frac{\mathrm{tr}\left[A_r^\top M\left(\xi,\boldsymbol{\theta}^0\right)A_r\right]+O\left(\|A_r\|_F^3\right)}{|||A_r|||_P^2}\right|\left[1+|||A_r|||_P^2K\right] \\
&= \lim_{r\to 0}\left|\frac{\mathrm{tr}\left[A_r^\top M\left(\xi,\boldsymbol{\theta}^0\right)A_r\right]}{|||A_r|||_P^2}+\frac{O\left(r^3\right)}{|||A_r|||_P^2}\right|\left[1+O\left(r^2\right)K\right] \\
&= \mathrm{tr}\left[A_P^\top M\left(\xi,\boldsymbol{\theta}^0\right)A_P\right] = \phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right),
\end{aligned}
$$

since $\frac{O(r^3)}{|||A_r|||_P^2}$ is $O(r)$ for $r\to 0$. On the other hand, one sees:

$$
\begin{aligned}
&\lim_{r\to 0} \inf_{\mathbf{T}=(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\mathscr{B}^k(\boldsymbol{\theta}^0,r)} \sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{|||\mathbf{T}_0-\mathbf{T}|||_P^2}+K\right] \\
&= \lim_{r\to 0}\inf_{A\in\mathcal{A}_r}\left|\frac{\mathrm{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right]+O\left(\|A\|_F^3\right)}{|||A|||_P^2}\right|\left[1+|||A|||_P^2K\right] \\
&= \lim_{r\to 0}\inf_{A\in\mathcal{A}_r}\left|\frac{\mathrm{tr}\left[\left(\frac{A}{\|A\|_F}\right)^\top M\left(\xi,\boldsymbol{\theta}^0\right)\left(\frac{A}{\|A\|_F}\right)\right]+O(r)}{\left|\left|\left|\left(\frac{A}{\|A\|_F}\right)\right|\right|\right|_P^2}\right|\left[1+|||A|||_P^2K\right] \\
&\geq \lim_{r\to 0}\inf_{A\in\mathcal{A}_r}\frac{\mathrm{tr}\left[\left(\frac{A}{\|A\|_F}\right)^\top M\left(\xi,\boldsymbol{\theta}^0\right)\left(\frac{A}{\|A\|_F}\right)\right]-Qr}{\left|\left|\left|\left(\frac{A}{\|A\|_F}\right)\right|\right|\right|_P^2} \\
&\geq \lim_{r\to 0}\inf_{A\in\mathbb{R}^{m\times k}}\frac{\mathrm{tr}\left[\left(\frac{A}{\|A\|_F}\right)^\top M\left(\xi,\boldsymbol{\theta}^0\right)\left(\frac{A}{\|A\|_F}\right)\right]-Qr}{\left|\left|\left|\left(\frac{A}{\|A\|_F}\right)\right|\right|\right|_P^2} \\
&= \lim_{r\to 0}\inf_{A:\|A\|_F=1}\frac{\mathrm{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right]-Qr}{|||A|||_P^2} \\
&= \lim_{r\to 0}\inf_{A:\|A\|_F=1}\frac{1-\frac{Qr}{\mathrm{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right]}}{\frac{|||A|||_P^2}{\mathrm{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right]}} \\
&\geq \lim_{r\to 0}\inf_{A:\|A\|_F=1}\frac{1-\frac{Qr}{\lambda}}{\frac{|||A|||_P^2}{\mathrm{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right]}} \\
&= \lim_{r\to 0}\left(1-\frac{Qr}{\lambda}\right)\inf_{A:\|A\|_F=1}\frac{1}{\frac{|||A|||_P^2}{\mathrm{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right]}} = \phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right),
\end{aligned}
$$

where we used the notation $\lambda = \frac{1}{k}\lambda_{\min}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right] > 0$ and that each matrix $A : \|A\|_F = 1$ includes at least one column $\mathbf{a}$ such that $1 \geq \|\mathbf{a}\|_{\ell(2)} \geq 1/\sqrt{k}$ and hence using (1.3)

$$\operatorname{tr}\left[A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\right] \geq \mathbf{a}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{a} = \|\mathbf{a}\|^2_{\ell(2)}\frac{\mathbf{a}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{a}}{\mathbf{a}^\top\mathbf{a}} \geq \frac{1}{k}\lambda_{\min}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right] = \lambda.$$

Moreover, there always exists such $r > 0$ that the expression $\left(1 - \frac{Qr}{\lambda}\right)$ is positive.

□

Theorem 4.18. *Suppose that the assumptions of Lemma 4.12 hold. Let* $M\left(\xi,\boldsymbol{\theta}^0\right) = \sum_{\mathbf{x}\in\mathcal{X}} M\left(\mathbf{x},\boldsymbol{\theta}^0\right)\xi(\mathbf{x})$ *be nonsingular, and* $\rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right) = |||\mathbf{T}_0 - \mathbf{T}|||_P$, *where* $\mathbf{T} = \left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right) \in \mathbb{R}^{m\times k}$, $\mathbf{T}_0 = \left(\boldsymbol{\theta}^0,\ldots,\boldsymbol{\theta}^0\right) \in \mathbb{R}^{m\times k}$ *and* $|||\cdot|||_P$ *is an arbitrary pseudonorm on* $\mathbb{R}^{m\times k}$. *Moreover, suppose that there is a matrix* $A_P \in \mathbb{R}^{m\times k}$ *such that* $|||A_P|||_P = 1$ *and* $\operatorname{tr}\left[A_P^\top M\left(\xi,\boldsymbol{\theta}^0\right)A_P\right] = \phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right) < \infty$. *Assume that there is* $\delta > 0$ *such that* $\mathscr{B}\left(\boldsymbol{\theta}^0,\delta\right) \subseteq \Theta$, *where* $\mathscr{B}\left(\boldsymbol{\theta}^0,\delta\right) = \left\{\boldsymbol{\theta} : \left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{\ell(2)} \leq \delta\right\}$ *and* Θ *denotes the parametric space. Suppose that there is no overlapping in* $\boldsymbol{\theta}^0$ *under the design* ξ, *i.e.* $\forall r > 0 \exists c_r > 0 \forall\, \boldsymbol{\theta} \in \Theta : \left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{\ell(2)} \geq r : \sum_{\mathbf{x}\in\mathcal{X}} I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi(\mathbf{x}) \geq c_r$, *then*

$$\lim_{K\to\infty}\inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)} + K\right] = \phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right).$$

Proof. According to Theorem 4.17 one has for any $K \geq 0$

$$\inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)} + K\right] \leq \phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right).$$

Moreover, the function

$$\inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)} + K\right]$$

is evidently non-decreasing in K, hence there is the limit for $K \to \infty$ and is less than or equal to $\phi_{|||\cdot|||_P}\left(\xi,\boldsymbol{\theta}^0\right)$.

Now we prove the opposite inequality. Since there is no overlapping, for every $r > 0$

$$\begin{aligned}
&\inf_{\substack{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k \\ \left\|\boldsymbol{\theta}^i-\boldsymbol{\theta}^0\right\|_{\ell(2)}>r\forall i=1,\ldots k}}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)} + K\right] \\
&\geq \inf_{\substack{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k \\ \left\|\boldsymbol{\theta}^i-\boldsymbol{\theta}^0\right\|_{\ell(2)}>r\forall i=1,\ldots k}}\sum_{i=1}^{k} 2c_r\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)} + K\right] \qquad (4.25) \\
&\geq 2kc_rK \xrightarrow[K\to\infty]{} \infty.
\end{aligned}$$

Hence for any $0 < r < \delta$,

$$\begin{aligned}
&\lim_{K\to\infty} \inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\Theta^k} \sum_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x}) \left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right] \\
&= \lim_{K\to\infty} \inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\mathscr{B}^k(\boldsymbol{\theta}^0,r)} \sum_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x}) \left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right],
\end{aligned} \tag{4.26}$$

since, as follows from (4.25), there is K_0 such that the infimum over the set

$$\left[\mathscr{B}^k(\boldsymbol{\theta}^0,r)\right]^{\mathsf{c}} = \left\{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k : \left\|\boldsymbol{\theta}^i-\boldsymbol{\theta}^0\right\|_{\ell(2)} > r\forall i=1,\ldots k\right\}$$

would be greater than $\phi_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right)$ for any $K > K_0$, while the infimum in (4.26) is less than or equal to $\phi_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right)$ for any K.

Finally, it follows from Theorem 4.17, that $\forall\epsilon>0\,\exists r_0\,\forall r<r_0$

$$\inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\mathscr{B}^k(\boldsymbol{\theta}^0,r)} \sum_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x}) \left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right] \geq \phi_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right)-\epsilon.$$

Hence, for every $\forall\epsilon>0\,\exists r_0\,\forall r<r_0$

$$\begin{aligned}
&\lim_{K\to\infty} \inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\mathscr{B}^k(\boldsymbol{\theta}^0,r)} \sum_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x}) \left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right] \\
&\geq \phi_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right)-\epsilon,
\end{aligned}$$

which together with (4.26) proves the theorem. □

So, to summarize, the extended pseudonorm criterion $\phi^{ext}_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right)$ coincides with the classical pseudonorm criterion $\phi_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right)$ if the model is normal linear regression model (Theorem 4.16). It approximately coincides with $\phi_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right)$ if the parametric space is a ball centred at $\boldsymbol{\theta}^0$ with a very small diameter (Theorem 4.17) or if the tuning constant K is very large (Theorem 4.18). **These theorems affirm that the criterion $\phi^{ext}_{\|\cdot\|_P}$ is an extension of the classical pseudonorm criterion $\phi_{\|\cdot\|_P}$.**

4.4.1 The case when $k = 1$

Let us summarize some well-known optimality criteria which are special cases of pseudonorm criteria. We will also formulate their extensions to the generalized regression models based on exponential families of distributions. First, we will consider the case $k = 1$.

The E-optimality criterion and its extension

One obtains the extended E-optimality criterion by taking the Euclidean norm $\|\cdot\|_{\ell(2)}$ as $|||\cdot|||_P$, since as follows from (1.3) and Def. 1.2 that $\phi_{\|\cdot\|_{\ell(2)}}\left(\xi,\boldsymbol{\theta}^0\right) = \min_{\mathbf{u}\in\mathbb{R}^m:\|\mathbf{u}\|_{\ell(2)}=1} \mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u} = \lambda_{\min}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right] = \phi_E\left(\xi,\boldsymbol{\theta}^0\right)$, hence

$$\phi_E^{ext}\left(\xi,\boldsymbol{\theta}^0\right) = \phi_{\|\cdot\|_{\ell(2)}}^{ext}\left(\xi,\boldsymbol{\theta}^0\right) = \inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}\right\|_{\ell(2)}^2}+K\right]. \tag{4.27}$$

The MV-optimality criterion and its extension

The convex MV-optimality criterion is defined as the maximal diagonal element of the matrix $M^{-1}\left(\xi,\boldsymbol{\theta}^0\right)$, which can be equivalently written as $\max_{\mathbf{u}\in\mathbb{R}^m:\|\mathbf{u}\|_{\ell(1)}=1} \mathbf{u}^\top M^{-1}\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u}$ (see López-Fidalgo et al., 1998) and is supposed to be minimized with respect to ξ. Using (4.4) and the fact that the $\ell(\infty)$ norm, $\|\mathbf{u}\|_{\ell(\infty)} = \max_{i=1,\ldots,m}|u_i|$, is dual to the $\ell(1)$ norm we can define the concave MV-criterion

$$\phi_{MV}\left(\xi,\boldsymbol{\theta}^0\right) = \phi_{\|\cdot\|_{\ell(\infty)}}\left(\xi,\boldsymbol{\theta}^0\right) = \min_{\mathbf{u}\in\mathbb{R}^m:\|\mathbf{u}\|_{\ell(\infty)}=1} \mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u},$$

which is supposed to be maximized with respect to ξ. Its extension is then given as

$$\begin{aligned}
\phi_{MV}^{ext}\left(\xi,\boldsymbol{\theta}^0\right) &= \phi_{\|\cdot\|_{\ell(\infty)}}^{ext}\left(\xi,\boldsymbol{\theta}^0\right) \\
&= \inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}\right\|_{\ell(\infty)}^2}+K\right] \\
&= \inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\max_{i=1,\ldots,m}\left(\theta_i^0-\theta_i\right)^2}+K\right].
\end{aligned}$$

The c-optimality criterion and its extension

Define the vector pseudonorm

$$|||\cdot|||_{P\mathbf{c}} : \mathbb{R}^m \mapsto [0,\infty),\ \mathbf{u}\to|||\mathbf{u}|||_{P\mathbf{c}} = \left|\mathbf{u}^\top\mathbf{c}\right|. \tag{4.28}$$

Lemma 5.6 from Pronzato and Pázman (2013) (or Theorem A1 in Appendix for $L=\mathbf{c}$) implies that

$$\begin{aligned}
\phi_{|||\cdot|||_{P\mathbf{c}}}\left(\xi,\boldsymbol{\theta}^0\right) &= \inf_{\mathbf{u}\in\mathbb{R}^m:|||\mathbf{u}|||_{P\mathbf{c}}=1} \mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u} \qquad = \inf_{\mathbf{u}\in\mathbb{R}^m:\left|\mathbf{u}^\top\mathbf{c}\right|=1} \mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u} \\
&= \begin{cases} \frac{1}{\mathbf{c}^\top M^-\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{c}} & \text{if } \mathbf{c}\in\mathcal{C}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right] \\ 0 & \text{otherwise} \end{cases} \quad = \phi_{\mathbf{c}}\left(\xi,\boldsymbol{\theta}^0\right),
\end{aligned}$$

see Def. 1.2. The extension is then

$$\phi^{ext}_{\|\|\cdot\|\|_{P\mathbf{c}}}\left(\xi,\boldsymbol{\theta}^0\right)=\inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\left\|\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}\right\|\right\|^2_{P\mathbf{c}}}+K\right]$$
$$=\inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left\{\frac{1}{\left[\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}\right)^\top\mathbf{c}\right]^2}+K\right\}.$$

We refer the reader also to a more general definition of the extended **c**-optimality criterion in Sect. 4.5.

The G-optimality criterion and its extension

Define the vector pseudonorm

$$\|\|\cdot\|\|_{PG}:\mathbb{R}^m\mapsto[0,\infty):\mathbf{u}\to\|\|\mathbf{u}\|\|_{PG}=\max_{\mathbf{x}\in\mathcal{X}}\left|\mathbf{u}^\top\mathbf{f}\left(\mathbf{x}\right)\right|. \tag{4.29}$$

Under the assumption that $\forall\mathbf{x}\in\mathcal{X}\,\mathbf{f}\left(\mathbf{x}\right)\neq\mathbf{0}$, the application of Lemma 5.6 from Pronzato and Pázman (2013) (one may also use Theorem A1 for $L=\mathbf{f}\left(\mathbf{x}\right)$) gives

$$\min_{\mathbf{x}\in\mathcal{X}}\inf_{\mathbf{u}:\left|\mathbf{u}^\top\mathbf{f}(\mathbf{x})\right|=1}\mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u}=$$
$$=\begin{cases}0 & \text{if }\exists\mathbf{x}\in\mathcal{X}:\mathbf{f}\left(\mathbf{x}\right)\notin\mathcal{C}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right],\\ \min_{\mathbf{x}\in\mathcal{X}}\frac{1}{\mathbf{f}^\top(\mathbf{x})M^-\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{f}(\mathbf{x})} & \text{otherwise,}\end{cases}$$

where the left hand side equals

$$\min_{\mathbf{x}\in\mathcal{X}}\inf_{\mathbf{u}\in\mathbb{R}^m:\mathbf{u}^\top\mathbf{f}(\mathbf{x})\neq0}\frac{\mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u}}{\left[\mathbf{u}^\top\mathbf{f}\left(\mathbf{x}\right)\right]^2}=\inf_{\mathbf{u}\in\mathbb{R}^m:\mathbf{u}^\top\mathbf{f}(\mathbf{x})\neq0}\frac{\mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u}}{\left[\max_{\mathbf{x}\in\mathcal{X}}\left|\mathbf{u}^\top\mathbf{f}\left(\mathbf{x}\right)\right|\right]^2}=$$
$$=\inf_{\mathbf{u}\in\mathbb{R}^m:\|\|\mathbf{u}\|\|_{PG}=1}\mathbf{u}^\top M\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{u}=\phi_{\|\|\cdot\|\|_{PG}}\left(\xi,\boldsymbol{\theta}^0\right).$$

It is easy to verify that $\|\|\cdot\|\|_{PG}$ is a norm if the set $\{\mathbf{f}\left(\mathbf{x}\right),\mathbf{x}\in\mathcal{X}\}$ includes m linearly independent vectors. Then, using Lemma 4.15, we have (see Def. 1.2)

$$\phi_{\|\|\cdot\|\|_{PG}}\left(\xi,\boldsymbol{\theta}^0\right)=\begin{cases}\min_{\mathbf{x}\in\mathcal{X}}\frac{1}{\mathbf{f}^\top(\mathbf{x})M^{-1}\left(\xi,\boldsymbol{\theta}^0\right)\mathbf{f}(\mathbf{x})} & \text{if } M(\xi,\boldsymbol{\theta}^0)\text{ is nonsingular}\\ 0 & \text{otherwise}\end{cases}=\phi_G\left(\xi,\boldsymbol{\theta}^0\right).$$

Moreover, even if $\|\|\cdot\|\|_{PG}$ is not norm, $\phi_{\|\|\cdot\|\|_{PG}}(\xi,\boldsymbol{\theta}^0)=\phi_G(\xi,\boldsymbol{\theta}^0)$ for $M(\xi,\boldsymbol{\theta}^0)$ nonsingular.

The extended $\phi_{\|\cdot\|_{PG}}$ optimality criterion is defined as

$$\phi^{ext}_{\|\cdot\|_{PG}}\left(\xi,\boldsymbol{\theta}^0\right) = \inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}\right\|^2_{P\mathbf{G}}}+K\right]$$

$$= \inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left\{\frac{1}{\max_{\mathbf{x}\in\mathcal{X}}\left[\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}\right)^\top\mathbf{f}\left(\mathbf{x}\right)\right]^2}+K\right\}.$$

Notice that the above mentioned criteria of **c**-, and G-optimality can be further generalized, see Sect. 4.5.

4.4.2 The case when $k=m$ and the distance $\rho\left(\cdot;\cdot,\cdots,\cdot\right)$ is given through Schatten p norm

One obtains different extended optimality criteria applying the Schatten p norm (4.1) as the distance ρ:

$$\inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta^m}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{m} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\left\|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^m\right)\right\|^2_{\mathrm{S}(p)}}+K\right]. \tag{4.30}$$

Lemma 4.19. *Let $U\in\mathbb{R}^{m\times m}$ be an orthogonal matrix and let $A\in\mathbb{R}^{m\times m}$. Then $\|AU\|_{\mathrm{S}(p)}=\|A\|_{\mathrm{S}(p)}$, where $\|\cdot\|_{\mathrm{S}(p)}$ is the Schatten p-norm.*

Proof. $\|AU\|_{\mathrm{S}(p)}=\left\{\mathrm{tr}\left[\left(AUU^\top A^\top\right)^{p/2}\right]\right\}^{1/p}=\left\{\mathrm{tr}\left[\left(AA^\top\right)^{p/2}\right]\right\}^{1/p}=\|A\|_{\mathrm{S}(p)}.$ □

Consequence 4.20. The Schatten norm of a matrix $\mathbf{T}$ is invariant to rearrangements of the columns of the matrix $\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\rightarrow\mathbf{T}P=\left(\boldsymbol{\theta}^{p_1},\ldots,\boldsymbol{\theta}^{p_m}\right)$, where P is some permutation matrix and $p_1,\ldots,p_m$ is the corresponding permutation of the numbers $1,\ldots,m$. Indeed, $\|\mathbf{T}P\|_{\mathrm{S}(p)}=\|\mathbf{T}\|_{\mathrm{S}(p)}$ since every permutation matrix is orthogonal (see e.g. Harville, 2008, Sect. 8.4c). Moreover, summation is a commutative operation and therefore in (4.30) we are searching for an m-tuple of vectors from Θ and the order of those vectors is not important.

The L_p-optimality criteria and their extension

We consider the class of concave L_p-criteria (see e.g. Kiefer, 1974, Eq. 4.11 or Kiefer, 1975, Eq. 2.1 for convex L_p-criteria or Pronzato and Pázman, 2013, Eq. 5.15 for their concave versions), but we restrict ourselves to $p\geq 1$,

$$\phi_{L_p}\left(\xi,\boldsymbol{\theta}\right)=\begin{cases}\frac{1}{\left(\mathrm{tr}\left\{\left[M^{-1}\left(\xi,\boldsymbol{\theta}\right)\right]^p\right\}\right)^{1/p}}=\frac{1}{\left\|M^{-1}\left(\xi,\boldsymbol{\theta}\right)\right\|_{\mathrm{S}(p)}} & \text{if } M\left(\xi,\boldsymbol{\theta}\right) \text{ is nonsingular,}\\ 0 & \text{otherwise.}\end{cases}$$

For instance, one obtains the A-optimality criterion if $p=1$ (directly from Def. 1.2) and the E-optimality criterion if $p\to\infty$ (see Kiefer, 1975, Eq. 2.1).

Lemma 4.21. *(Dette et al., 1995, p. 36) Let $q=2p/(p+1)$, $p\geq 1$. Then $\phi_{L_p}(\xi,\boldsymbol{\theta})=\phi_{\|\cdot\|_{\mathrm{S}(q)}}(\xi,\boldsymbol{\theta})$.*

Proof. We follow the proof given in Dette et al. (1995) on p. 36. We will use the notation $M=M(\xi,\boldsymbol{\theta})$ since the statement of the lemma does depend on $M(\xi,\boldsymbol{\theta})$, not on ξ and $\boldsymbol{\theta}$ separately. The case when M is singular is proved in Lemma 4.15, hence we proceed by nonsingular M.

According to (4.3), it suffices to prove that

$$\max_{A\in\mathbb{R}^{m\times m}:\|A\|^D_{\mathrm{S}(q)}=1}\mathrm{tr}\left[A^\top M^{-1}A\right]=\left\|M^{-1}\right\|_{\mathrm{S}(p)}.$$

Since $\|\cdot\|_{\mathrm{S}\left(\frac{q}{q-1}\right)}$ is the dual norm to $\|\cdot\|_{\mathrm{S}(q)}$, we obtain

$$\begin{aligned}
\max_{A\in\mathbb{R}^{m\times m}:\|A\|^D_{\mathrm{S}(q)}=1}\mathrm{tr}\left[A^\top M^{-1}A\right] &= \max_{A\in\mathbb{R}^{m\times m}:\|A\|_{\mathrm{S}\left(\frac{q}{q-1}\right)}=1}\mathrm{tr}\left[A^\top M^{-1}A\right]\\
&= \max_{A\in\mathbb{R}^{m\times m}:\|A\|_{\mathrm{S}\left(\frac{2p}{p-1}\right)}=1}\mathrm{tr}\left[A^\top M^{-1}A\right]\\
&= \max_{A\in\mathbb{R}^{m\times m}:\left\|AA^\top\right\|_{\mathrm{S}\left(\frac{p}{p-1}\right)}=1}\mathrm{tr}\left[M^{-1}AA^\top\right]\\
&= \max_{\substack{F\in\mathbb{R}^{m\times m}:\|F\|_{\mathrm{S}\left(\frac{p}{p-1}\right)}=1\\ F\text{ is positive semidefinite}}}\mathrm{tr}\left[M^{-1}F\right],
\end{aligned}$$

where we used for $r=\frac{2p}{p-1}\geq 2$ that if $\|A\|_{\mathrm{S}(r)}=1$, then

$$1=\|A\|^2_{\mathrm{S}(r)}=\left\{\mathrm{tr}\left[\left(AA^\top\right)^{r/2}\right]\right\}^{2/r}=\left\{\mathrm{tr}\left[\left(AA^\top AA^\top\right)^{r/4}\right]\right\}^{2/r}=\left\|AA^\top\right\|_{\mathrm{S}\left(\frac{r}{2}\right)}.$$

Then, using Def. 4.4, we have

$$\begin{aligned}
&\max_{\substack{F\in\mathbb{R}^{m\times m}:\|F\|_{\mathrm{S}\left(\frac{p}{p-1}\right)}=1\\ F\text{ is positive semidefinite}}}\mathrm{tr}\left[M^{-1}F\right]\leq\\
&\qquad\leq\max_{\substack{F\in\mathbb{R}^{m\times m}\\ \|F\|_{\mathrm{S}\left(\frac{p}{p-1}\right)}=1}}\left|\mathrm{tr}\left[M^{-1}F\right]\right|=\left\|M^{-1}\right\|^D_{\mathrm{S}\left(\frac{p}{p-1}\right)}=\left\|M^{-1}\right\|_{\mathrm{S}(p)}.
\end{aligned}$$

Now, to prove the opposite inequality, we set

$$A'=\frac{M^{-(p-1)/2}}{\left\|M^{-(p-1)/2}\right\|_{\mathrm{S}\left(\frac{2p}{p-1}\right)}}=\frac{M^{-(p-1)/2}}{\left\{\mathrm{tr}\left[M^{-p}\right]\right\}^{(p-1)/2p}}.$$

One sees $\|A'\|_{\mathrm{S}\left(\frac{2p}{p-1}\right)}=1$ and hence we obtain

$$\begin{aligned}\max_{A\in\mathbb{R}^{m\times m}:\|A\|^D_{\mathrm{S}(q)}=1}\mathrm{tr}\left[A^\top M^{-1}A\right] &= \max_{A\in\mathbb{R}^{m\times m}:\|A\|_{\mathrm{S}\left(\frac{2p}{p-1}\right)}=1}\mathrm{tr}\left[A^\top M^{-1}A\right]\\ &\geq \mathrm{tr}\left[A'^\top M^{-1}A'\right]=\frac{\mathrm{tr}\left[M^{-p}\right]}{\left\{\mathrm{tr}\left[M^{-p}\right]\right\}^{(p-1)/p}}\\ &=\left[\mathrm{tr}\left(M^{-p}\right)\right]^{1-(p-1)/p}=\left\{\mathrm{tr}\left[\left(M^{-1}M^{-1}\right)^{p/2}\right]\right\}^{1/p}\\ &=\left\|M^{-1}\right\|_{\mathrm{S}(p)}.\end{aligned}$$

□

As a consequence of Lemma 4.21, the L_p-criteria with $1\leq p=\frac{q}{2-q}$, $1\leq q\leq 2$ are only special cases of pseudonorm criteria based on the Schatten norm, and hence it is easy to derive their extended versions, see (4.30). Particularly, when the distance ρ from Def. 4.14 is chosen as $\rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)=\left\|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^m\right)\right\|_{\mathrm{S}(1)}$, one obtains the extension of A-optimality

$$\phi_A^{ext}=\inf_{\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta^m}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\|\mathbf{T}_0-\mathbf{T}\|^2_{\mathrm{S}(1)}}+K\right],$$

and for $\rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)=\left\|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^m\right)\right\|_{\mathrm{S}(2)}=\left\|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^m\right)\right\|_F$ one has the extension of E-optimality criterion

$$\phi_E^{ext'}=\inf_{\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta^m}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\|\mathbf{T}_0-\mathbf{T}\|^2_{F}}+K\right],$$

where in both cases we used the notation $\mathbf{T}_0=\left(\boldsymbol{\theta}^0,\ldots,\boldsymbol{\theta}^0\right)$.

In Sect. 4.4.1 we defined the extended E-optimality ϕ_E^{ext} via the Euclidean norm. Theorems 4.17 and 4.16 indicate that the extensions ϕ_E^{ext} and $\phi_E^{ext'}$ approximately coincide when the parametric space Θ is a ball in $\mathbb{R}^m$ centred at $\boldsymbol{\theta}^0$ with a very small diameter or they coincide if the model is linear. Generally, we managed to prove that $\phi_E^{ext}\geq\phi_E^{ext'}\geq\frac{1}{m}\phi_E^{ext}$. We have from the definition of the Frobenius norm (4.2) that

$$\begin{aligned}&\phi_E^{ext'}\left(\xi,\boldsymbol{\theta}^0\right)\\ &=\inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta^m}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{m}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\left\|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^m\right)\right\|^2_F}+K\right]\\ &=\inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta^m}\sum_{i=1}^{m}\sum_{\mathbf{x}\in\mathcal{X}}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\sum_{i=1}^m\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|^2_{\ell(2)}}+K\right].\end{aligned}$$

It follows that

$$\begin{aligned}
&\phi_E^{ext'}\left(\xi,\boldsymbol{\theta}^0\right)\\
&\leq \inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta\times\left\{\boldsymbol{\theta}^0\right\}\times\ldots\times\left\{\boldsymbol{\theta}^0\right\}} \sum_{i=1}^{m}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\sum_{i=1}^{m}\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|_{\ell(2)}^2}+K\right]\\
&= \inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}\right\|_{\ell(2)}^2}+K\right]=\phi_E^{ext}\left(\xi,\boldsymbol{\theta}^0\right).
\end{aligned}$$

On the other hand, without loss of generality, one may assume that $\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^1\right\|\geq\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^i\right\|\ \forall i=1,\ldots,m$ (see Consequence 4.20) and hence

$$\begin{aligned}
&\phi_E^{ext'}\left(\xi,\boldsymbol{\theta}^0\right)\\
&\geq \inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta^m}\sum_{i=1}^{m}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\sum_{i=1}^{m}\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^1\right\|_{\ell(2)}^2}+K\right]\\
&\geq \inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^m\right)\in\Theta^m}\sum_{i=1}^{m}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{m\left\|\boldsymbol{\theta}^0-\boldsymbol{\theta}^1\right\|_{\ell(2)}^2}+\frac{K}{m}\right]=\frac{1}{m}\phi_E^{ext}\left(\xi,\boldsymbol{\theta}^0\right).
\end{aligned}$$

4.4.3 The case when $k<m$ and the distance $\rho\left(\cdot;\cdot,\cdots,\cdot\right)$ is given through a pseudonorm

The L-optimality criterion and its extension

Assume that $L\in\mathbb{R}^{m\times k}$ $k<m$ is a given nonzero matrix. The L-optimality criterion is defined as

$$\phi_L\left(\xi,\boldsymbol{\theta}^0\right)=\begin{cases}\frac{1}{\operatorname{tr}\left(L^\top M^-\left(\xi,\boldsymbol{\theta}^0\right)L\right)} & \text{if } \mathcal{C}\left(L\right)\subseteq\mathcal{C}\left[M\left(\xi,\boldsymbol{\theta}^0\right)\right],\\ 0 & \text{otherwise,}\end{cases}$$

where $M^-\left(\xi,\boldsymbol{\theta}^0\right)$ is an arbitrary generalized inverse of $M\left(\xi,\boldsymbol{\theta}^0\right)$ (see e.g. Pronzato and Pázman (2013) p. 116).

Define now the pseudonorm

$$\left|\!\left|\!\left|\cdot\right|\!\right|\!\right|_{PL}:\mathbb{R}^{m\times k}\mapsto[0,\infty):A\to\left|\!\left|\!\left|A\right|\!\right|\!\right|_{PL}=\left|\operatorname{tr}\left(A^\top L\right)\right|. \tag{4.31}$$

Then it follows from Theorem A1 that

$$\begin{aligned}
\phi_{\left|\!\left|\!\left|\cdot\right|\!\right|\!\right|_{PL}}\left(\xi,\boldsymbol{\theta}^0\right)&=\inf_{A\in\mathbb{R}^{m\times k}:\left|\!\left|\!\left|A\right|\!\right|\!\right|_{PL}=1} A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A=\inf_{A\in\mathbb{R}^{m\times k}:\left|\operatorname{tr}\left(A^\top L\right)\right|=1} A^\top M\left(\xi,\boldsymbol{\theta}^0\right)A\\
&=\phi_L\left(\xi,\boldsymbol{\theta}^0\right),
\end{aligned}$$

and the extended L-optimality criterion is

$$\phi_L^{ext}\left(\xi,\boldsymbol{\theta}^0\right)=\phi_{|||\cdot|||_{PL}}^{ext}=\inf_{\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{|||\mathbf{T}-\mathbf{T}_0|||_{PL}^2}+K\right]$$
$$=\inf_{\mathbf{T}=\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k}\sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\mathrm{tr}^2\left[\left(\mathbf{T}-\mathbf{T}_0\right)^\top L\right]}+K\right],$$

where $\mathbf{T}_0=\left(\boldsymbol{\theta}^0,\ldots,\boldsymbol{\theta}^0\right)$.

The members of the class of pseudonorm optimality criteria (4.22), which were presented in Sects. 4.4.1–4.4.3, are summarized in Table 4.1. The required extended optimality criterion (4.23) is then obtained directly by applying the corresponding pseudonorm.

optimality criterion	k	corresponding pseudonorm
E	1	$\|\cdot\|_{\ell(2)}$
MV	1	$\|\cdot\|_{\ell(\infty)}$
$\mathbf{c}$	1	$\|\|\|\cdot\|\|\|_{P\mathbf{c}}$ as defined in (4.28)
G	1	$\|\|\|\cdot\|\|\|_{PG}$ as defined in (4.29)
L_p, $p\geq 1$	m	$\|\cdot\|_{\mathrm{S}(q)}$ with $q=2p/(p+1)$
A	m	$\|\cdot\|_{\mathrm{S}(1)}$
L	$m>k>1$	$\|\|\|\cdot\|\|\|_{PL}$ as defined in (4.31)

Table 4.1: Some well-known optimality criteria involved in the class of pseudonorm criteria (4.22). The parameter k and the corresponding pseudonorm are also indicated.

4.5 Alternative extensions of c- and G-optimality criteria in generalized regression models

Definition 4.22. Let $K\geq 0$ be a tuning constant. In a generalized regression model based on exponential families of distributions we define

the extended c-optimality criterion as

$$\phi_{\mathbf{c}}^{ext}\left(\xi,\boldsymbol{\theta}^0\right)=\inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\left[\frac{1}{\left|h\left(\boldsymbol{\theta}^0\right)-h\left(\boldsymbol{\theta}\right)\right|^2}+K\right]\xi\left(\mathbf{x}\right),$$

where $h:\Theta\to\mathbb{R}$ is a given function of parameters, and

the extended G-optimality criterion as

$$\phi_G^{ext}\left(\xi,\boldsymbol{\theta}^0\right)=\inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}}2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\left[\frac{1}{\max_{\mathbf{x}\in\mathcal{X}}\left[\alpha\left(\mathbf{x},\boldsymbol{\theta}^0\right)-\alpha\left(\mathbf{x},\boldsymbol{\theta}\right)\right]^2}+K\right]\xi\left(\mathbf{x}\right),$$

where $\mathbf{x}\in\mathcal{X}\rightarrow\alpha\left(\mathbf{x},\boldsymbol{\theta}\right)$ is a given regression function (especially, one may take $\alpha\left(\mathbf{x},\boldsymbol{\theta}\right)=g\left(\mathbf{x},\boldsymbol{\theta}\right)$ or $\alpha\left(\mathbf{x},\boldsymbol{\theta}\right)=\mu\left(\mathbf{x},\boldsymbol{\theta}\right)$).

The criteria from Def. 4.22 were together with the extended E-optimality criterion (4.27) studied in our paper Burclová and Pázman (2016b). In the nonlinear regression model (1.5) with normally distributed random errors ε with zero mean and unit variance, these criteria coincide with the extended criteria from Pázman and Pronzato (2014), since then $2I_{\mathbf{x}}(\boldsymbol{\theta}^0,\boldsymbol{\theta})=\left[\eta\left(\mathbf{x},\boldsymbol{\theta}^0\right)-\eta(\mathbf{x},\boldsymbol{\theta})\right]^2$, see Remark 4.10.

One sees that when $h\left(\boldsymbol{\theta}\right)=\boldsymbol{\theta}^\top\mathbf{c}$ for some $\mathbf{c}\in\mathbb{R}^m\setminus\{\mathbf{0}\}$, then the criterion $\phi_{\mathbf{c}}^{ext}\left(\xi,\boldsymbol{\theta}^0\right)$ coincides with the criterion $\phi_{\|\cdot\|_{P\mathbf{c}}}^{ext}\left(\xi,\boldsymbol{\theta}^0\right)$ from Sect. 4.4.1 since $\left|h\left(\boldsymbol{\theta}^0\right)-h(\boldsymbol{\theta})\right|^2=\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{P\mathbf{c}}^2$.

Similarly, when $\alpha\left(\mathbf{x},\boldsymbol{\theta}\right)=\boldsymbol{\theta}^\top\mathbf{f}\left(\mathbf{x}\right)$ with $\mathbf{f}\left(\mathbf{x}\right)\in\mathbb{R}^m\setminus\{\mathbf{0}\}\ \forall\mathbf{x}\in\mathcal{X}$, then the criterion $\phi_G^{ext}\left(\xi,\boldsymbol{\theta}^0\right)$ corresponds to the extended pseudonorm criterion $\phi_{\|\cdot\|_{PG}}^{ext}\left(\xi,\boldsymbol{\theta}^0\right)$ from Sect. 4.4.1, since $\max_{\mathbf{x}\in\mathcal{X}}\left[\alpha\left(\mathbf{x},\boldsymbol{\theta}^0\right)-\alpha\left(\mathbf{x},\boldsymbol{\theta}\right)\right]^2=\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\right\|_{PG}^2$.

Provided that the criteria $\phi_{\mathbf{c}}^{ext}$ and ϕ_G^{ext} coincide with the corresponding extended pseudonorm criteria (as indicated above), in the normal linear regression model (1.4) with random errors of zero mean and unit variance, $\phi_{\mathbf{c}}^{ext}\left(\xi,\boldsymbol{\theta}^0\right)=\phi_{\mathbf{c}}\left(\xi,\boldsymbol{\theta}^0\right)$ and $\phi_G^{ext}\left(\xi,\boldsymbol{\theta}^0\right)=\phi_{\|\cdot\|_{PG}}\left(\xi,\boldsymbol{\theta}^0\right)\ \forall\boldsymbol{\theta}^0\in\Theta=\mathbb{R}^m\ \forall\xi\in\Xi$, where we directly applied Theorem 4.16. Notice that the criterion $\phi_{\|\cdot\|_{PG}}$ coincides with classical G-optimality if $\|\cdot\|_{PG}$ is norm or if $M\left(\xi,\boldsymbol{\theta}^0\right)$ is nonsingular, see Sect. 4.4.1.

4.6 Properties of extended optimality criteria defined in (4.21)

One sees that the extended optimality criteria (4.21) are positively homogeneous and concave in ξ, since they are defined as infimum of functions linear in ξ.

The concavity ensures the existence of the directional derivative. Pázman and Pronzato (2014) in Theorems 2 and 4 or Pronzato and Pázman (2013) in Theorems 7.16–7.17 derived the directional derivatives for the extended E- and $\mathbf{c}$-optimality criteria in the nonlinear regression model (1.5). See also Theorem 3 from Pázman and Pronzato (2014) for the formulation of the equivalence theorem.

The extended optimality criteria (4.21) are local in the sense that the nominal parameter value $\boldsymbol{\theta}^0$ is required. On the other hand, they are global since they are

taking into account the parameter values distant from $\boldsymbol{\theta}^0$. The undesirable dependence on $\boldsymbol{\theta}^0$ can be avoided by considering $\min_{\boldsymbol{\theta}^0 \in \Theta} \phi_\rho^{ext}\left(\xi, \boldsymbol{\theta}^0\right)$, as suggested in Pázman and Pronzato (2014).

4.6.1 Parameter K

The tuning parameter K in (4.21) plays very important role. By an adequate choice of K we can express our preference for the precision or for the stability of the estimator.

For $K = 0$, the infimum in (4.21) may be reached simply because $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k$ are very distant points from $\boldsymbol{\theta}^0$ (i.e. $\rho\left(\boldsymbol{\theta}^0; \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k\right)$ is very large) even if the I-divergence $\sum_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^k I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}^i\right) \xi(\mathbf{x})$ is not necessarily small, and hence there is no real instability in the model.

On the other hand, as $K \to \infty$ we obtain the classical (not extended) pseudonorm optimality criteria (see Theorem 4.18), and hence the extended criteria lose their global properties for very large K.

4.7 Algorithm and examples

The design $\xi^\star$ is $\phi_\rho^{ext}\left(\cdot, \boldsymbol{\theta}^0\right)$ optimal if $\xi^\star = \arg\max_{\xi \in \Xi} \phi_\rho^{ext}\left(\xi, \boldsymbol{\theta}^0\right)$. We can use LP methods to compute $\xi^\star$, since the extended criteria (4.21) are defined as infimum of functions linear in ξ:

$$\phi_\rho^{ext}\left(\xi, \boldsymbol{\theta}^0\right) = \inf_{\left(\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k\right) \in \Theta^k} \sum_{\mathbf{x} \in \mathcal{X}} H_k\left(\mathbf{x}, \boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k\right) \xi(\mathbf{x}),$$

where

$$H_k\left(\mathbf{x}, \boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k\right) = \sum_{i=1}^k 2 I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}^i\right) \left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0; \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k\right)} + K\right]. \tag{4.32}$$

The following algorithm and examples were presented in Burclová and Pázman (2016b).

Example 4.23. The example illustrates the possible instabilities related to the maximum likelihood estimation of parameter $\boldsymbol{\theta}$ in a generalized regression model. We were inspired by the Example 1 in Pázman and Pronzato (2014). Consider now the binomial model with the pmf

$$f(y, \mathbf{x}, \theta) = \binom{n}{y} p(\mathbf{x}, \theta)^y \left[1 - p(\mathbf{x}, \theta)\right]^{n-y}, \tag{4.33}$$

which can be rewritten in the exponential form (4.15):

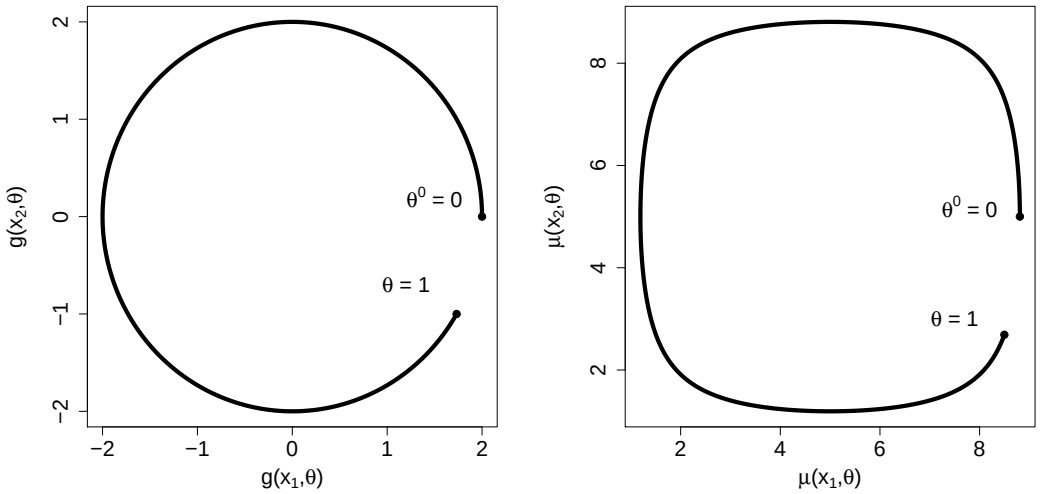
$$f(y, \mathbf{x}, \theta) = \exp\left\{\ln \binom{n}{y} + y g(\mathbf{x}, \theta) - n \ln\left[1 + e^{g(\mathbf{x}, \theta)}\right]\right\}, \tag{4.34}$$

with $g(\mathbf{x},\theta) = \ln\{p(\mathbf{x},\theta)/[1-p(\mathbf{x},\theta)]\}$, and with the mean of sufficient statistic $t(y) = y$ equal to $\mu(\mathbf{x},\theta) = np(\mathbf{x},\theta) = ne^{g(\mathbf{x},\theta)}/\left\{1+e^{g(\mathbf{x},\theta)}\right\}$ (the logistic function, see also Appendix B.1). We set $n = 10$ and apply the regression function (similar to that in the Example 1 of Pázman and Pronzato, 2014)

$$g(\mathbf{x},\theta) = 2\cos(t-u\theta);\ \mathbf{x} = (t,u)^\top. \tag{4.35}$$

The experiment consists of two trials, one in $\mathbf{x}_1 = (0,u)^\top$ and the second in $\mathbf{x}_2 = (\pi/2,u)^\top$. The design problem is to find the optimal value $u \in \left[0, \frac{11}{6}\pi\right]$ for the maximum likelihood estimation of the unknown parameter $\theta \in [0,1]$. The information matrix $M_u(\theta) \equiv M(\mathbf{x}_1,\theta) + M(\mathbf{x}_2,\theta)$ is obtained from (4.18). Suppose that the true and nominal parameter values equal 0, i.e. $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}^0 = 0$. Hence, the classical design approach, which maximizes the information matrix $M_u\left(\theta^0\right) = nu^2$, leads to the locally optimal design achieved at $u_{loc} = \frac{11}{6}\pi$.

In Figure 4.1 are the circular canonical surface $\mathscr{K} = \left\{(g(\mathbf{x}_1,\theta), g(\mathbf{x}_2,\theta))^\top;\ \theta \in [0,1]\right\}$ and the expectation surface (which is not circular due to the nonlinearity of the logistic function) $\mathscr{E} = \left\{(\mu(\mathbf{x}_1,\theta), \mu(\mathbf{x}_2,\theta))^\top;\ \theta \in [0,1]\right\}$ depicted for the design u_{loc}. Since the surfaces are nearly overlapping, the maximum likelihood estimate $\hat{\theta}_2$ (see (4.14)) can be, with a large probability, in the neighbourhood of $\theta = 1$, which indicates the possibility of a false maximum likelihood estimate.



(a) Canonical surface (b) Expectation surface

Figure 4.1: Example 4.23: The canonical and expectation surface for locally optimal design $u_{loc} = \frac{11\pi}{6}$.

Now we can use the extended criterion in the following form

$$\phi^{ext}\left(u,\theta^0\right) = \min_{\theta\in[0,1]} I\left(\theta^0,\theta;u\right)/\left(\theta-\theta^0\right)^2$$

with the I-divergence $I\left(\theta^0,\theta;u\right)\equiv I_{\mathbf{x}_1}\left(\theta^0,\theta\right)+I_{\mathbf{x}_2}\left(\theta^0,\theta\right)$, where in the binomial model one has

$$I_{\mathbf{x}}\left(\theta^0,\theta\right)=n\left[p\left(\mathbf{x},\theta^0\right)\ln\frac{p\left(\mathbf{x},\theta^0\right)}{p\left(\mathbf{x},\theta\right)}+\left\{1-p\left(\mathbf{x},\theta^0\right)\right\}\ln\left\{\frac{1-p\left(\mathbf{x},\theta^0\right)}{1-p\left(\mathbf{x},\theta\right)}\right\}\right] \tag{4.36}$$

(see Appendix B.1). The numerical optimization indicates that the criterion $\phi^{ext}\left(u,\theta^0\right)$ is maximized approximately at $u_{ext}=\pi$, and for this value is the maximum likelihood estimator more stable and the probability of a false $\hat{\theta}_2$ is negligible. As seen even in Figure 4.2, the points $\left(g\left(\mathbf{x}_1,\theta\right),g\left(\mathbf{x}_2,\theta\right)\right)^\top$ for $\theta=0$ and $\theta=1$ are now as distant as possible. The same holds for $\left(\mu\left(\mathbf{x}_1,\theta\right),\mu\left(\mathbf{x}_2,\theta\right)\right)^\top$.

We could also include the tuning parameter K and consider the extended criterion in the form $\min_{\theta\in[0,1]} I\left(\theta^0,\theta;u\right)\left[\frac{1}{(\theta-\theta^0)}+K\right]^2$. The optimal u would be between π and $\frac{11}{6}\pi$ for K positive.

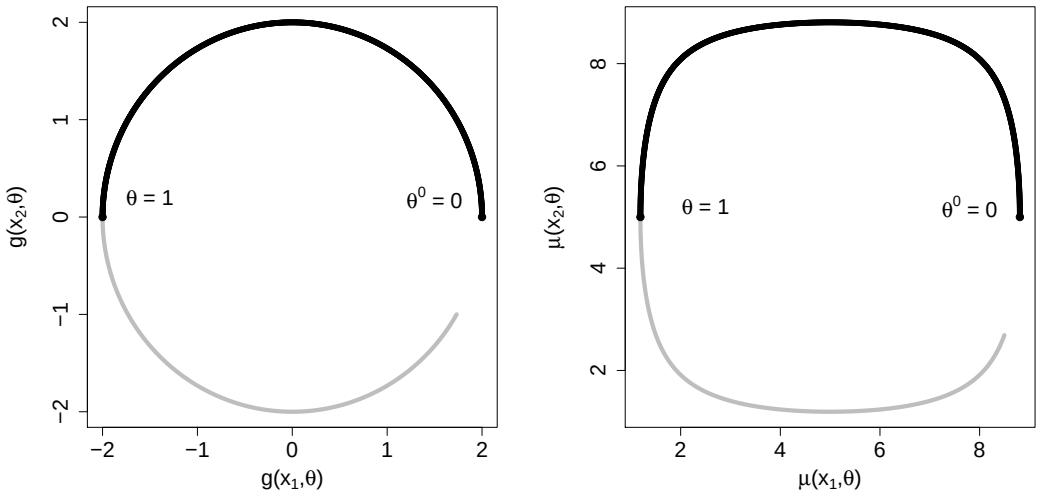
The dependence of $H_u\left(\theta^0,\theta\right)=I\left(\theta^0,\theta;u\right)/\left(\theta-\theta^0\right)^2$ on θ for different values of u is displayed in Figure 4.3. One sees that $u_{loc}=\frac{11\pi}{6}$ maximizes $H_u\left(\theta^0,\theta\right)$ for θ near the nominal value θ^0, i.e. u_{loc} is optimal in the local sense. On the other hand, minimum of $H_u\left(\theta^0,\theta\right)$ over θ is maximized for $u_{ext}=\pi$, i.e. u_{ext} is optimal in the global sense. Finally, $u=\pi/4$ is optimal neither from the global nor from the local point of view.

We performed a simulation where we 10000 times repeated the experiment (4.34)–(4.35) consisting of two measurements $\mathbf{x}_1=(0,u)^\top$ and $\mathbf{x}_2=(\pi/2,u)^\top$ for $u_{loc}=\frac{11\pi}{6}$ and $u_{ext}=\pi$, respectively. The true parameter value $\bar{\theta}$ was equal to 0. In the case of u_{loc}, the simulated probability of a false maximum likelihood estimate, i.e. the probability that $\hat{\theta}_2>0.5$, was approximately 0.1981, for u_{ext} the same probability was approximately equal to 0.0021, which is significantly smaller. In Figure 4.4 are displayed some likelihoods $L(\theta)=f\left(y_1,\mathbf{x}_1,\theta\right)f\left(y_2,\mathbf{x}_2,\theta\right)$ as functions of θ. One sees that for u_{ext} is the likelihood usually maximized near $\bar{\theta}=0$, which is not true for u_{loc}. △

In more difficult situations than in Example 4.23, one may use iterative Algorithm 4.24, which was suggested in Pázman and Pronzato (2014). Algorithm 4.24 optimizes the extended criterion (4.21) reformulated in (4.32), for $k=1$, i.e. we will consider $H_1\left(\mathbf{x},\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)=2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}\right)}+K\right]$. The modification for $k>1$ would be straightforward. We remind that in this chapter we suppose that $\mathcal{X}$ is finite and hence the design ξ is a vector from $\mathbb{R}^{\mathrm{card}(\mathcal{X})}$.

Algorithm 4.24.

0. i Choose the starting design $\xi^{(0)}\in\mathbb{R}^{\mathrm{card}(\mathcal{X})}$, s.t. $\xi^{(0)}(\mathbf{x})\geq 0\ \forall\mathbf{x}\in\mathcal{X}$ and $\sum_{\mathbf{x}\in\mathcal{X}}\xi^{(0)}(\mathbf{x})=1$.



(a) Canonical surface (b) Expectation surface

Figure 4.2: Example 4.23: The canonical and expectation surface for $u_{ext} = \pi$ (black color, gray color corresponds to u_{loc}) which maximizes the criterion $\phi^{ext}\left(u, \boldsymbol{\theta}^0\right)$.

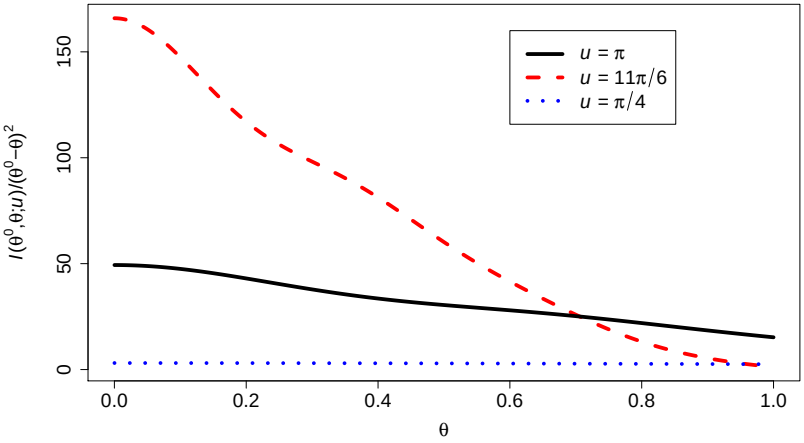


Figure 4.3: Example 4.23: $H_u(\theta^0, \theta)$ as a function of θ for different u.

ii Take ϵ greater than 0 but small.

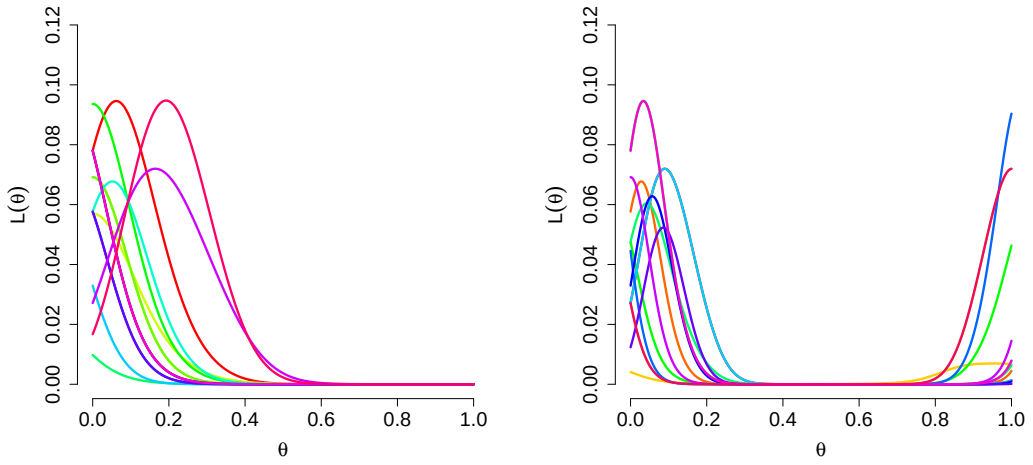
iii Set $\Theta^{(0)} = \emptyset$.

iv Construct a finite set $\mathcal{G}^{(0)}$ in Θ.

v Set $i = 0$.

1. Compute $\boldsymbol{\theta}^{(i+1)} = \arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{\mathbf{x} \in \mathcal{X}} H_1\left(\mathbf{x}, \boldsymbol{\theta}^0, \boldsymbol{\theta}\right) \xi^{(i)}(\mathbf{x})$ as follows:

   i Compute $\tilde{\boldsymbol{\theta}}^{(i+1)} = \arg\min_{\boldsymbol{\theta} \in \mathcal{G}^{(i)}} \sum_{\mathbf{x} \in \mathcal{X}} H_1\left(\mathbf{x}, \boldsymbol{\theta}^0, \boldsymbol{\theta}\right) \xi^{(i)}(\mathbf{x})$.



(a) $u_{loc} = \frac{11\pi}{6}$ (b) $u_{ext} = \pi$

Figure 4.4: Some likelihoods based on the simulation of the experiment in Example 4.23.

ii Perform a local minimization over Θ initialized at $\tilde{\boldsymbol{\theta}}^{(i+1)}$ and denote by $\boldsymbol{\theta}^{(i+1)}$ the solution.

2. Set $\mathcal{G}^{(i+1)} = \mathcal{G}^{(i)} \cup \left\{\boldsymbol{\theta}^{(i+1)}\right\}$.
3. Set $\Theta^{(i+1)} = \Theta^{(i)} \cup \left\{\boldsymbol{\theta}^{(i+1)}\right\}$.
4. Use an LP solver to find $\left(t^{(i+1)}, \xi^{(i+1)}\right)$ which maximizes $t^{(i+1)}$ under the constraints:
   - $t^{(i+1)} > 0$,
   - $\xi^{(i+1)}(\mathbf{x}) \geq 0 \; \forall \, \mathbf{x} \in \mathcal{X}$,
   - $\sum_{\mathbf{x} \in \mathcal{X}} \xi^{(i+1)}(\mathbf{x}) = 1$,
   - $\sum_{\mathbf{x} \in \mathcal{X}} H_1\left(\mathbf{x}, \boldsymbol{\theta}^0, \boldsymbol{\theta}\right) \xi^{(i+1)}(\mathbf{x}) \geq t^{(i+1)} \; \forall \boldsymbol{\theta} \in \Theta^{(i+1)}$.
5. Set $\Delta^{(i+1)} = t^{(i+1)} - \phi_\rho^{ext}\left(\xi^{(i+1)}, \boldsymbol{\theta}^0\right)$.
6. If $\Delta^{(i+1)} < \epsilon$, take $\xi^{(i+1)}$ as an ϵ-optimal design and stop, else set $i \leftarrow i+1$ and continue from Step 1.

Notice that when the parametric space Θ is finite, then we can obtain the optimal $\xi^\star$ after first iteration as a solution of the LP problem formulated in Step 4 (the inequalities in the LP problem then have to be satisfied for any $\boldsymbol{\theta} \in \Theta$, not $\Theta^{(i+1)}$).

The stopping rule in Step 5 follows from following inequalities:

$$\phi_\rho^{ext}\left(\xi^{(i+1)},\boldsymbol{\theta}^0\right) \le \max_{\xi\in\Xi}\phi_\rho^{ext}\left(\xi,\boldsymbol{\theta}^0\right) = \max_{\xi\in\Xi}\inf_{\boldsymbol{\theta}\in\Theta}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi(\mathbf{x})\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}\right)}+K\right] \le$$

$$\le \max_{\xi\in\Xi}\inf_{\boldsymbol{\theta}\in\Theta^{(i+1)}}\sum_{\mathbf{x}\in\mathcal{X}} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi(\mathbf{x})\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}\right)}+K\right] = t^{(i+1)}.$$

We can compute the criterion value $\phi_\rho^{ext}\left(\xi^{(i+1)},\boldsymbol{\theta}^0\right)$ in Step 5 of the algorithm similarly as the optimization in Step 1, i.e.

i Compute $\tilde{\boldsymbol{\theta}}^{(i+2)} = \arg\min_{\boldsymbol{\theta}\in\mathcal{G}^{(i+1)}}\sum_{\mathbf{x}\in\mathcal{X}} H_1\left(\mathbf{x},\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi^{(i+1)}(\mathbf{x})$.

ii Perform a local minimization over Θ initialized at $\tilde{\boldsymbol{\theta}}^{(i+2)}$ and denote by $\boldsymbol{\theta}^{(i+2)}$ the solution.

iii $\phi_\rho^{ext}\left(\xi^{(i+1)},\boldsymbol{\theta}^0\right) = \sum_{\mathbf{x}\in\mathcal{X}} H_1\left(\mathbf{x},\boldsymbol{\theta}^0,\boldsymbol{\theta}^{(i+2)}\right)\xi^{(i+1)}(\mathbf{x})$.

If the algorithm continues to the next iteration, we can skip Step 1 and directly use the result $\boldsymbol{\theta}^{(i+2)}$ of the above procedure.

The computations in the next example were performed in Matlab computing environment and we used the `linprog` function with the default interior point algorithm to solve the LP problems.

Example 4.25. In this example we would like to present a numerical computation of the extended E-optimum design (see Eq. (4.27)), using Algorithm 4.24. The mean of the random variable y observed at the design point $\mathbf{x} = (x_1, x_2)^\top$ was chosen according to Pázman and Pronzato (2014), Example 2:

$$\mu(\mathbf{x},\boldsymbol{\theta}) = np(\mathbf{x},\boldsymbol{\theta}) = \frac{n}{6}\left\{1+\theta_1 x_1+\theta_1^3(1-x_1)+\theta_2 x_2+\theta_2^2(1-x_2)\right\},\ \boldsymbol{\theta} = (\theta_1,\theta_2)^\top. \tag{4.37}$$

Here we again consider a binomial model with y distributed according to (4.33), with $n = 10$ and $p(\mathbf{x},\boldsymbol{\theta})$ given by (4.37).

We maximized the extended E-optimality criterion (4.27) with the binomial I-divergence (4.36) and the nominal parameter value $\boldsymbol{\theta}^0 = (1/8, 1/8)^\top$ from the parametric space $\Theta = [-1,1]\times[0,2]$. The design space $\mathcal{X} = \{0, 0.1, \ldots, 0.9, 1\}^2$ was finite.

The set $\mathcal{G}^{(0)}$ corresponded to a random Latin hypercube design with 10000 points renormalized to Θ. The starting design $\xi^{(0)}$ was taken as uniform measure on $\mathcal{X}$.

For the accuracy $\epsilon = 10^{-10}$, the algorithm stopped after 20 iterations for $K = 0$ and after 47 iterations for $K = 5$. The results are given in Table 4.2. △

K	$\xi^\star$	$\phi_E^{ext}\left(\xi^\star, \boldsymbol{\theta}^0\right)$ for $K=0$	$\phi_E^{ext}\left(\xi^\star, \boldsymbol{\theta}^0\right)$ for $K=5$
0	$\begin{Bmatrix} (0,0)^\top & (0,1)^\top & (1,1)^\top \\ 0.345 & 0.029 & 0.626 \end{Bmatrix}$	0.0215	0.0249
5	$\begin{Bmatrix} (0,0)^\top & (1,0)^\top & (0,1)^\top & (1,1)^\top \\ 0.247 & 0.072 & 0.197 & 0.484 \end{Bmatrix}$	0.0165	0.1972

Table 4.2: Example 4.25: ϕ_E^{ext}-optimal designs for $K=0$ and $K=5$ computed by Algorithm 4.24 and corresponding values of ϕ_E^{ext}, see (4.27), for $K=0$ and $K=5$.

4.8 Conclusions of new results in this chapter

In Sect. 4.2 we summarized some ways of designing experiments in generalized regression models taking into account the results of previous chapters of this thesis.

The rest of the chapter was devoted to the extended optimality criteria, the main purpose of which is to avoid the possible instabilities appearing when estimating parameters in regression models.

While Pázman and Pronzato (2014), who introduced the extended optimality criteria, considered the instabilities related to the least squares estimation in classical nonlinear regression (1.5), we considered the maximum likelihood estimation in generalized regression models. The potential instabilities were described in Sect. 4.3.

The first and most important outcome of this chapter is that in Sects. 4.3, 4.4, and 4.5 we formulated the extended criteria related to the maximum likelihood estimation. We restricted ourselves to the generalized regression models based on exponential families of distributions. The reason of this restriction is that in the exponential family is the I-divergence (see Def. 4.9) markedly simplified (the integral is removed, see Eq. (4.11)) and can be approximated by the Fisher information matrix, see Lemma 4.12. Hence the extended criteria defined by us are based on the I-divergence, which is related to the information obtained in the experiment and, simultaneously, helps to identify the instabilities in the model.

The second important outcome is that we formulated the extended version of the criterion of MV-optimality (Sect. 4.4.1), A-optimality (Sect. 4.4.2), and L-optimality (Sect. 4.4.3). This is a new result also for classical nonlinear regression, since Pázman and Pronzato (2014) defined only the extensions of E-, $\mathbf{c}$-, and G-optimality criteria. We remark that the extended A- and L-optimality require the minimization over a k-tuple of points $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^k$, $k > 1$. The relations of all considered extended criteria (E, $\mathbf{c}$, G, MV, A, L) to their classical versions were presented in Theorems 4.16, 4.17, and 4.18.

Section 4.6 provides some properties of extended optimality criteria and Sect. 4.7 consists of one illustrative example and of one example, where the optimal design was computed iteratively via the Algorithm suggested in Pázman and Pronzato (2014).

Conclusion

The thesis is focused on designing experiments in nonlinear models. The usual aim of such experiments is to obtain an accurate estimate of the model parameter. However, the presence of this unknown parameter in the information matrix and in the criterion function makes the optimization of the experiments in nonlinear models more complicated comparing to the same task in linear regression models. On the other hand, the nonlinear models, and especially generalized regression models, become very popular in many scientific and financial areas.

In the thesis we considered two different approaches to experimental design. The first (and more usual) approach uses the asymptotic properties of parameter estimators and focuses on maximizing the corresponding information matrix (or Fisher information matrix in the case of generalized regression models) and is applied in Chaps. 2, 3 and in Sect. 4.2.1.

Chapter 2 uses an LP method for calculating the local and average D-, A-, E-, and E_k-optimal designs following from proper reformulations of these criteria by methods of linear algebra. As the main contribution of this chapter we consider the method for computing the criterion robust design of Harman (2004), which is a quite difficult nondifferentiable problem that cannot be handled by many standard algorithms.

In Chap. 3 we follow the considerations of Pázman and Pronzato (2007) that the local, AVE, and maximin criteria do not always satisfactory reflect the whole parametric space. They suggested to use the quantile criterion as an alternative, but they admitted that its concavity is not ensured. From this point of view, the application of the CVaR criterion—used in Valenzuela et al. (2015) for the first time—seems to be more attractive. We prove that the CVaR criterion is a compromise between the AVE and maximin criterion, and after deriving the subgradient we apply the cutting plane method to obtain CVaR optimal designs. We formulate the equivalence theorem and some other relevant results based on the risk theory. We think that also the examples in this chapter are interesting, especially the last one, where we are able to compare AVE, maximin, CVaR, and quantile optimal designs. The question of approaching

quantile optimal design by using CVaR criterion remains still open.

The asymptotic approach to optimal experimental design appears also in Sect. 4.2.1, where the application of results of Chaps. 2 and 3 to generalized regression models is indicated.

The remainder of Chap. 4 is based on the second (different) approach that prevents unstable maximum likelihood estimates in generalized regression models. Hence, instead of maximizing the information matrix, here we maximize the suitably standardized I-divergence and so we extend the results of Pázman and Pronzato (2014). In particular, Pázman and Pronzato (2014) pointed out the problem of stability and uniqueness of the estimator in classical nonlinear model and formulated the criteria of extended E-, $\mathbf{c}$-, and G-optimality. Here we reformulate those criteria for purposes of maximum likelihood estimation in generalized regression models and, moreover, we define the extended criteria of MV-, A- and L-optimality that can also be applied in classical nonlinear regression. We prove that the extended optimality criteria coincide with the classical optimality criteria if the model is linear with normally distributed random errors or in the case of a suitably restricted parametric space. The danger of possibly false maximum likelihood estimate is demonstrated on the illustrative example and the numerical example applies the algorithm for maximizing the extended optimality criteria as suggested in Pázman and Pronzato (2014).

Resumé

Úvod

Dizertačná práca sa zaoberá optimalizovaním experimentov v nelineárnych regresných modeloch, najmä novými kritériami optimality vhodnými pre nelineárne modely.

Experiment nazveme optimálnym pokiaľ prináša experimentátorovi najviac informácie o neznámom parametri. Štandardne sa maximalizuje tzv. informačná matica, ktorá síce „meria“ veľkosť tejto informácie, ale len lokálne pretože závisí práve od hodnoty neznámeho parametra. V práci sme sa venovali aj rozšíreniu alternatívneho prístupu uvedeného v článku Pázman a Pronzato (2014), kde informačná matica nezohráva tak kľúčovú úlohu.

Navrhovanie experimentov v nelineárnom regresnom modeli

Nech meranie $y(\mathbf{x})$ spĺňa nelineárny regresný model, teda $y(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon$, kde $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^m$ je neznámy parameter a $\eta : \mathcal{X} \times \Theta \mapsto \mathbb{R}$ je spojité diferencovateľné zobrazenie na kompaktnom parametrickom priestore Θ. Predpokladajme, že $\mathcal{X}$ je konečná množina bodov $\mathbf{x}$, v ktorých môžeme vykonať merania a že všetky merania v experimente sú vykonávané nezávisle. Návrh experimentu ξ je potom ľubovoľné pravdepodobnostné rozdelenie na $\mathcal{X}$ a Ξ je množina všetkých takých návrhov ξ. O náhodných chybách ε predpokladáme, že majú nulovú strednú hodnotu a konštantnú (neznámu) varianciu. Pre $\boldsymbol{\theta}^0 \in \mathrm{int}(\Theta)$ a návrh $\xi \in \Xi$ má informačná matica tvar

$$M\left(\xi, \boldsymbol{\theta}^0\right) = \sum_{\mathbf{x} \in \mathcal{X}} M\left(\mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x}),$$

kde elementárna informačná matica je

$$M\left(\mathbf{x}, \boldsymbol{\theta}^0\right) = \left.\frac{\partial \eta\left(\mathbf{x}, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \left.\frac{\partial \eta\left(\mathbf{x}, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}.$$

Cieľom štandardného prístupu k navrhovaniu experimentov je maximalizovať lokálne kritérium optimality, teda zvolenú reálnu funkciu $\phi\left(\xi, \boldsymbol{\theta}^0\right)$, ktorá meria veľkosť informačnej matice $M\left(\xi, \boldsymbol{\theta}^0\right)$ (viď napr. monografie Pronzato a Pázman (2013), Fedorov a

Leonov (2014)). V súčasnosti poznáme širokú škálu rôznych kritérií, medzi ktoré patria kritériá D-, A-, E-, $\mathbf{c}$- a G-optimality (viď napr. Kap. 5.1.2 v Pronzato a Pázman (2013)).

V nelineárnom modeli sa nevyhneme závislosti kriteriálnej funkcie na parametri $\boldsymbol{\theta}$. Pri navrhovaní experimentu teda buď uvažujeme nominálnu hodnotu parametra $\boldsymbol{\theta}^0$ nachádzajúcu sa v blízkosti skutočnej hodnoty (tzv. lokálne kritérium), alebo maximalizujeme kritérium $\phi(\xi, \boldsymbol{\theta})$ pri najhoršej možnej hodnote parametra (tzv. maximinné kritérium) či uvažujeme nejaké apriórne rozdelenie π na parametrickom priestore Θ a následne maximalizujeme strednú hodnotu $\phi(\xi, \boldsymbol{\theta})$ vzhľadom na toto apriórne rozdelenie (tzv. priemerovacie kritérium).

Navrhovanie experimentov v zovšeobecnenom regresnom modeli založenom na exponenciálnej triede rozdelení

V dizertačnej práci sme sa zamerali na zovšeobecnené regresné modely založené na exponenciálnej triede hustôt, teda pri danom $\mathbf{x} \in \mathcal{X}$ a neznámom parametri $\boldsymbol{\theta} \in \Theta$ pozorujeme $\mathbf{y} \in \mathbb{R}^l$ s hustotou

$$f(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) = \exp\left\{-\psi(\mathbf{y}) + t^\top(\mathbf{y}) g(\mathbf{x}, \boldsymbol{\theta}) - \kappa\left[g(\mathbf{x}, \boldsymbol{\theta})\right]\right\}, \tag{5.1}$$

kde ψ a κ sú známe funkcie, $t(\mathbf{y}) \in \mathbb{R}^r$ je postačujúca štatistika pre parameter $\boldsymbol{\theta}$ a $g(\mathbf{x}, \boldsymbol{\theta})$ zodpovedá tzv. kanonickému parametru. Modely založené na (5.1) zahŕňajú aj klasickú nelineárnu regresiu s normálne rozdelenými chybami či logistickú regresiu. Odhad parametra $\boldsymbol{\theta}$ sa počíta metódou maximálnej vierohodnosti. Modely popísané hustotou (5.1) boli uvažované napr. v Pázman (1993), ale v práci sme sa opierali o vlastnosti exponenciálnych tried rozdelení rozoberaných v Brown (1986); Barndorff-Nielsen (1978); Efron (1978). Z uvedenej literatúry vyplýva, že elementárna Fisherova informačná matica pre meranie z (5.1) má tvar

$$M(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial g^\top(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Sigma(\mathbf{x}, \boldsymbol{\theta}) \frac{\partial g(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top},$$

kde $\Sigma(\mathbf{x}, \boldsymbol{\theta})$ je kovariančná matica postačujúcej štatistiky $t(\mathbf{y})$ pri danom $\mathbf{x}$ a $\boldsymbol{\theta}$.

Klasický prístup navrhovania experimentov v zovšeobecnených regresných modeloch (nie nutne založených na exponenciálnych triedach rozdelení) bol popísaný napr. v Atkinson et al. (2014) a opiera sa o maximalizáciu vhodných funkcií matice $\sum_{\mathbf{x} \in \mathcal{X}} M(\mathbf{x}, \boldsymbol{\theta}) \xi(\mathbf{x})$ vzhľadom na ξ. Podobne ako v obyčajnej nelineárnej regresii je možné použiť lokálne, maximinné a priemerovacie kritériá na odstránenie nežiaducej závislosti od $\boldsymbol{\theta}$.

V práci sme však za účelom navrhovania experimentov použili aj I-divergenciu (Kullback-Leiblerovu I divergenciu, viď Kullback a Leibler (1951)), ktorá meria vzdialenosť medzi dvomi rozdeleniami pravdepodobnosti a ukázali sme jej istý súvis s Fisherovou iformačnou maticou. Pre naše potreby bolo nutné merať vzdialenosť dvoch hustôt (5.1) pri hodnote parametra $\boldsymbol{\theta}^0$ a $\boldsymbol{\theta}$. V modeloch založených na exponenciálnych triedach má táto I-divergencia tvar

$$I_{\mathbf{x}}\left(\boldsymbol{\theta}^0, \boldsymbol{\theta}\right) = \mu^{\top}\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\left[g\left(\mathbf{x}, \boldsymbol{\theta}^0\right) - g\left(\mathbf{x}, \boldsymbol{\theta}\right)\right] + \kappa\left[g\left(\mathbf{x}, \boldsymbol{\theta}\right)\right] - \kappa\left[g\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\right],$$

kde $\mu\left(\mathbf{x}, \boldsymbol{\theta}^0\right)$ označuje strednú hodnotu $t(\mathbf{y})$ pri $\mathbf{x}$ a $\boldsymbol{\theta}^0$.

Ciele dizertačnej práce

Dizertačnú prácu a jej ciele možno rozdeliť do troch viac-menej nezávislých celkov:

1. Navrhovanie experimentov pomocou lineárneho programovania:
   - prepísať kritériá D-, A- a E_k-optimality do takej formy, ktorá umožňuje využitie lineárneho programovania pri optimalizácii experimentu a
   - využiť tieto nové formulácie kritérií na riešenie zložitejších problémov ako je hľadanie tzv. robustného návrhu vzhľadom na triedu ortogonálne invariantných kritérií alebo optimalizovanie experimentu pri doplnkových lineárnych ohraničeniach.
2. Navrhovanie experimentov pomocou kritérií inšpirovaných teóriou rizika:
   - definovať kritérium založené na podmienenej hodnote v riziku ako konkávnu funkciu návrhu ξ a uvažovať aj diskrétne apriórne rozdelenia (vychádzajúc z Valenzuela et al. (2015), kde bolo toto kritérium použité za účelom navrhovania experimentu po prvýkrát),
   - analyzovať a interpretovať toto kritérium,
   - študovať vzťah tohto kritéria k priemerovaciemu, maximinnému, lokálnemu a kvantilovému kritériu,
   - odvodiť smerovú deriváciu a dokázať vetu o ekvivalencii pre toto kritérium a
   - ukázať možnosť výpočtu optimálnych návrhov pomocou lineárneho programovania.
3. Formulácia rozšírených kritériií optimality za účelom obmedzenia mylných odhadov v zovšeobecnených regresných modeloch:

- predefinovať rozšírené kritériá E-, $\mathbf{c}$- a G-optimality z Pázman a Pronzato (2014) tak, aby boli aplikovateľné aj v zovšeobecnených regresných modeloch založených na exponenciálnych triedach rozdelení,
- navrhnúť rozšírené verzie aj pre kritériá MV-, L- a A-optimality,
- dokázať, že tieto kritériá sú naozaj rozšírením klasických kritérií optimality a
- aplikovať lineárne programovanie a kritérium založené na podmienenej hodnote v riziku aj v zovšeobecnených regresných modeloch.

Výsledky práce

Navrhovanie experimentov pomocou lineárneho programovania

V Kap. 9.5.3 monografie Pronzato a Pázman (2013) bolo ukázané akým spôsobom možno použiť metódy lineárneho programovania na optimalizáciu experimentov pomocou kritérií E-, $\mathbf{c}$- a G-optimality.

V článku Burclová a Pázman (2016a), na ktorom je založená táto časť, sa podarilo pomocou maticovej algebry odvodiť ekvivalentné formulácie ďalších konkávnych a pozitívne homogénnych kritérií optimality vhodných pre aplikáciu lineárneho programovania, menovite:

kritérium D-optimality

$$\begin{aligned}\forall \xi \in \Xi^+ \ \phi_D(\xi, \boldsymbol{\theta}^0) &\equiv \left\{\det\left[M(\xi, \boldsymbol{\theta}^0)\right]\right\}^{1/m} \\ &= \min_{\zeta \in \Xi^+} \sum_{\mathbf{x} \in \mathcal{X}} \frac{\det^{1/m}\left[M\left(\zeta, \boldsymbol{\theta}^0\right)\right]}{m} \mathbf{f}^\top\left(\mathbf{x}, \boldsymbol{\theta}^0\right) M^{-1}\left(\zeta, \boldsymbol{\theta}^0\right) \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x}),\end{aligned}$$

kde $\Xi^+ = \left\{\xi : M\left(\xi, \boldsymbol{\theta}^0\right) \text{ je regulárna}\right\}$ a $\mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right) = \frac{\partial \eta(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$,

kritérium A-optimality

$$\begin{aligned}\forall \xi \in \Xi^+ \ \phi_A(\xi, \boldsymbol{\theta}^0) &\equiv \frac{1}{\operatorname{tr}\left\{\left[M\left(\xi, \boldsymbol{\theta}^0\right)\right]^{-1}\right\}} \\ &= \min_{\zeta \in \Xi^+} \sum_{\mathbf{x} \in \mathcal{X}} \frac{\left\|M^{-1}\left(\zeta, \boldsymbol{\theta}^0\right) \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right)\right\|^2_{\ell(2)}}{\left\{\operatorname{tr}\left[M^{-1}\left(\zeta, \boldsymbol{\theta}^0\right)\right]\right\}^2} \xi(\mathbf{x}),\end{aligned}$$

kde $\|\cdot\|_{\ell(2)}$ označuje euklidovskú normu,

kritérium E_k-optimality

$$\forall \xi \in \Xi \; \phi_{E_k}(\xi, \boldsymbol{\theta}^0) \equiv \sum_{i=1}^{k} \lambda_i \left[M\left(\xi, \boldsymbol{\theta}^0\right) \right] = \min_{\zeta \in \Xi} \sum_{\mathbf{x} \in \mathcal{X}} \left\| P^{(k)}\left(\zeta, \boldsymbol{\theta}^0\right) \mathbf{f}\left(\mathbf{x}, \boldsymbol{\theta}^0\right) \right\|^2_{\ell(2)} \xi(\mathbf{x}),$$

kde $\lambda_1 \left[M\left(\xi, \boldsymbol{\theta}^0\right) \right] \leq \ldots \leq \lambda_m \left[M\left(\xi, \boldsymbol{\theta}^0\right) \right]$ sú usporiadané vlastné čísla matice $M\left(\xi, \boldsymbol{\theta}^0\right)$, $\mathbf{u}_1 \left[M\left(\xi, \boldsymbol{\theta}^0\right) \right], \ldots, \mathbf{u}_m \left[M\left(\zeta, \boldsymbol{\theta}^0\right) \right]$ sú zodpovedajúce ortonormálne vlastné vektory a $P^{(k)}\left(\zeta, \boldsymbol{\theta}^0\right) = \sum_{i=1}^{k} \mathbf{u}_i \left[M\left(\zeta, \boldsymbol{\theta}^0\right) \right] \mathbf{u}_i^\top \left[M\left(\zeta, \boldsymbol{\theta}^0\right) \right]$ je ortogonálny projektor.

Z uvedeného vyplýva, že tieto kritériá optimality sa dajú vyjadriť pomocou vhodne zvolenej funkcie $H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right)$ a množiny Ξ^* ako

$$\phi\left(\xi, \boldsymbol{\theta}^0\right) = \min_{\zeta \in \Xi^*} \sum_{\mathbf{x} \in \mathcal{X}} H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi(\mathbf{x}).$$

Optimálny návrh experimentu $\xi^\star$ maximalizujúci kritérium $\phi\left(\xi, \boldsymbol{\theta}^0\right)$ je potom riešením problému lineárneho programovania s nekonečne veľa lineárnymi ohraničeniami:

$$\begin{aligned} &\max\left(\mathbf{0}^\top, 1\right) \begin{pmatrix} \xi \\ t \end{pmatrix} \\ &\text{tak, že } \sum_{\mathbf{x} \in \mathcal{X}} H\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right) \xi\left(\mathbf{x}\right) \geq t \text{ pre každé } \zeta \in \Xi^*, \\ &\qquad \xi\left(\mathbf{x}\right) \geq 0 \text{ pre každé } \mathbf{x} \in \mathcal{X}, \text{ a } \sum_{\mathbf{x} \in \mathcal{X}} \xi\left(\mathbf{x}\right) = 1. \end{aligned} \tag{5.2}$$

Tento problém sme riešili pomocou iteračného algoritmu navrhnutého v Shimizu a Aiyoshi (1980). Neskôr sa ukázalo, že k rovnakému algoritmu a prepisu kritérií môžene dospieť použitím subgradientov (viď napr. Kap. 3.1.5–3.1.6 z Nesterov (2004)) a metódy *cutting planes* z článku Kelley (1960), viď. Kap. 9.5.3 z Pronzato a Pázman (2013). Použitý algoritmus poskytuje pravidlo zastavenia, ktoré sa líši od bežných pravidiel založených na vete o ekvivalencii.

Ďalej sme sa v práci venovali riešeniu komplexnejších problémov navrhovania experimentov, kde môžu byť uplatnené práve metódy lineárneho programovania. Jedným z nich je problém hľadania návrhu experimentu robustného vzhľadom na triedu tzv. ortogonálne invariantných kritérií. Harman (2004) dokázal, že takýto návrh je riešením nasledovnej optimalizačnej úlohy

$$\xi^\star_{\text{ef}} = \arg\max_{\xi \in \Xi} \min_{1 \leq k \leq m} \left[\frac{\phi_{E_k}\left(\xi, \boldsymbol{\theta}^0\right)}{\max_{\nu \in \Xi} \phi_{E_k}\left(\nu, \boldsymbol{\theta}^0\right)} \right]. \tag{5.3}$$

Úlohu (5.3) je možné formulovať ako úlohu lineárneho programovania s nekonečne veľa ohraničeniami a opäť aplikovať algoritmus z Shimizu and Aiyoshi (1980). Hlavnou

myšlienkou je, že najprv vypočítame hodnotu menovateľa vo výraze v (5.3) pre každé k (napríklad metódou lineárneho programovania) a označíme ju $E_k\left(opt, \boldsymbol{\theta}^0\right)$. Problém lineárneho programovania, ktorý rieši (5.3) potom pozostáva aj z týchto ohraničení:

$$\sum_{\mathbf{x}\in\mathcal{X}} \frac{H_{E_k}\left(\zeta, \mathbf{x}, \boldsymbol{\theta}^0\right)}{E_k\left(opt, \boldsymbol{\theta}^0\right)} \xi(\mathbf{x}) \geq t \text{ pre každé } \zeta \in \Xi \text{ a pre každé } k \in \{1, \ldots, m\},$$

kde H_{E_k} označuje funkciu H zodpovedajúcu kritériu ϕ_{E_k}. Podotýkame, že v článku Filová et al. (2012) bol problém (5.3) riešený metódami semidefinitného programovania.

Všimnime si, že medzi lineárne ohraničenia v (5.2) ľahko pridáme ďalšie — napríklad lineárne ohraničenie na cenu experimentu. Dodaním ďalších lineárnych ohraničení, ako sme ukázali v Burclová a Pázman (2016a), môžeme tiež optimalizovať jedno kritérium pri súčasnom dolnom ohraničení na hodnotu iného kritéria.

Ďalej sme dokázali vetu, ktorá rozširuje naše úvahy z lokálnych na priemerovacie kritériá v nelineárnych regresných modeloch.

Veta *Nech π je apriórne rozdelenie na parametrickom priestore Θ. Označme $\Xi_\Theta = \{\xi : M(\xi, \boldsymbol{\theta})$ je regulárna $\forall \boldsymbol{\theta} \in \Theta\}$. Môžeme písať*

$$\int_\Theta \phi(\xi, \boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}) = \min_{\zeta\in\Xi^*} \sum_{\mathbf{x}\in\mathcal{X}} H_{AVE}(\zeta, \mathbf{x})\, \xi(\mathbf{x}),$$

pre každé $\xi \in \Xi^*$, kde $H_{AVE}(\zeta, \mathbf{x}) = \int_\Theta H(\zeta, \mathbf{x}, \boldsymbol{\theta})\, d\pi(\boldsymbol{\theta})$ a $\Xi^* = \Xi_\Theta$ pre D- a A-optimalitu a $\Xi^* = \Xi$ pre kritériá E_k-optimality.

V dizertačnej práci sme sa venovali aj možnosti ako tieto výsledky rozšíriť pre prípad navrhovania experimentov v zovšeobecnených regresných modeloch založených na exponenciálnej triede rozdelení. Hlavný rozdiel pri odvodzovaní preformulovaných kritérií D-, A- a E_k-optimality spočíva v použití matice $F(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial g^\top(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Sigma^{1/2}(\mathbf{x}, \boldsymbol{\theta})$ namiesto vektora $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$.

Navrhovanie experimentov pomocou kritérií inšpirovaných teóriou rizika

Pázman a Pronzato (2007) poukazujú vo svojej práci na niektoré nedostatky priemerovacích a maximinných kritérií, ktoré sú bežne používané na odstránenie závislosti kritéria optimality od skutočnej (neznámej) hodnoty parametra. Ako alternatívu navrhujú použiť kvantilové kritérium pri pevne zvolenej hladine $\alpha \in [0, 1]$

$$\Phi_\alpha^Q(\xi) = \max\{t \in \mathbb{R} : Pr[\phi(\xi, \boldsymbol{\theta}) \geq t] \geq 1 - \alpha\}.$$

Spomínané nedostatky sa síce kvantilovým kritériom odstránili, ale zároveň autori pripúšťajú výraznú nevýhodu kvantilového kritéria, ktorou je jeho nekonkávnosť, resp. nekonvexnosť, čo so sebou prináša isté obtiažnosti pri výpočte optimálnych návrhov.

Na druhej strane Valenzuela et al. (2015) poprvýkrát navrhli použitie podmienenej hodnoty v riziku (CVaR, z ang. *Conditional Value at Risk*) pri navrhovaní experimentov a inšpirovali sa pri tom prácami z oblasti teórie rizika. Na základe prác Pflug (2000), Rockafellar a Uryasev (2000), Rockafellar a Uryasev (2002) sa nám podarilo previesť hlbšiu analýzu kritéria založeného na CVaR. Hlavnou výhodou tohto kritéria, ktorá bola zmienená už v článku Valenzuela et al. (2015), je jeho konkávnosť (aby sme boli úplne presní, práce Pflug (2000), Rockafellar a Uryasev (2000), Rockafellar a Uryasev (2002), Valenzuela et al. (2015) uvažujú konvexnú verziu CVaR, čo je typické pre teóriu rizika) a súčasne, že má veľmi podobné vlastnosti ako kvantilové kritérium.

Nami definované CVaR kritérium je pre $\alpha \in (0,1]$ dané nasledovným predpisom

$$\Phi_\alpha(\xi) = \max_{c\in\mathbb{R}} \left\{ c + \frac{1}{\alpha} E\left[\min\{0, \phi(\xi,\boldsymbol{\theta}) - c\}\right] \right\}. \tag{5.4}$$

Kritérium je konkávne v ξ, ak je aj pôvodné kritérium $\phi(\xi,\boldsymbol{\theta})$ konkávne v ξ a jeho formulácia v (5.4) sa najviac podobá na definíciu CVaR v článku Pflug (2000), ale vychádza zo vzťahu prvotne odvodeného v Rockafellar a Uryasev (2000). Na základe článku Rockafellar a Uryasev (2002) vieme, že jedným z bodov, ktoré pre dané ξ riešia maximalizačný problém v (5.4) je bod $c = \Phi_\alpha^Q(\xi)$. Vychádzajúc z Rockafellar a Uryasev (2002), kde je dokázané ekvivalentné tvrdenie, sme ukázali, že CVaR kritérium sa nachádza vždy medzi dvomi podmienenými strednými hodnotami, konkrétne

$$E[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) < \Phi_\alpha^Q(\xi)] \le \Phi_\alpha(\xi) \le E[\phi(\xi,\boldsymbol{\theta}) \mid \phi(\xi,\boldsymbol{\theta}) \le \Phi_\alpha^Q(\xi)], \tag{5.5}$$

čo nám pomohlo interpretovať CVaR kritérium (5.4). Uvedené výsledky platia pre ľubovoľné apriórne rozdelenie π na parametrickom priestore Θ, a teda pre ľubovoľnú náhodnú premennú $\phi(\xi,\boldsymbol{\theta})$ (diskrétnu alebo spojitú). Ak je $\phi(\xi,\boldsymbol{\theta})$ spojitá náhodná premenná, tak (5.5) platí so znamienkom rovnosti a navyše

$$\Phi_\alpha(\xi) = \frac{1}{\alpha} \int_{\left\{\boldsymbol{\theta}:\phi(\xi,\boldsymbol{\theta}) \le \Phi_\alpha^Q(\xi)\right\}} \phi(\xi,\boldsymbol{\theta})\, d\pi(\boldsymbol{\theta}),$$

čo je výraz analogický k definícii CVaR kritéria v článku Valenzuela et al. (2015).

V ďalšej časti sme sa viac zaoberali CVaR kritériom z pohľadu navrhovania experimentov. Podarilo sa nám ukázať, že pre $\alpha = 1$ sa CVaR kritérium zhoduje s priemerovacím kritériom a za určitých podmienok konverguje CVaR kritérium pre $\alpha \to 0$ k maximinnému kritériu. Kým priemerovacie kritérium môže viesť k veľmi zlým hodnotám pre niektoré body parametrického priestoru, maximinné kritérium je zas príliš orientované na okraje parametrického priestoru (viď Pázman a Pronzato (2007)). Práve vhodnou voľbou parametra α by sme mohli dosiahnuť primeraný kompromis medzi týmito dvomi kritériami, ktorý bude dostatočne robustný a súčasne nebude

zameraný len na okrajové body. Hodnota $\alpha = 0.5$ súvisí s mediánom a vedie k podobným výsledkom ako priemerovacie kritérium, preto zrejme vhodná voľba parametra α bude niekde z intervalu $(0, 0.5)$. Musíme podotknúť, že na rozdiel od kvantilového kritéria, kritérium CVaR nie je invariantné vzhľadom na nelineárne škálovanie pôvodného kritéria $\phi(\xi, \boldsymbol{\theta})$.

Keďže CVaR kritérium je konkávne, má zmysel preňho sformulovať vetu o ekvivalencii, ktorá úzko súvisí so smerovou deriváciou. Podarilo sa nám ukázať, že ak je kritérium $\phi(\xi, \boldsymbol{\theta})$ konkávna a spojitá funkcia v ξ pre každé $\boldsymbol{\theta} \in \Theta$, tak návrh $\xi^\star$ je CVaR optimálny práve vtedy ak

$$\sup_{\nu \in \Xi, b \in \mathbb{R}} \left[b - c + \frac{1}{\alpha} E \left(\begin{cases} 0 & \text{ak } \phi(\xi^\star, \boldsymbol{\theta}) > c \\ \min\{0, \mathcal{F}_{\phi(\cdot, \boldsymbol{\theta})}(\xi^\star, \nu) - (b - c)\} & \text{ak } \phi(\xi^\star, \boldsymbol{\theta}) = c \\ \mathcal{F}_{\phi(\cdot, \boldsymbol{\theta})}(\xi^\star, \nu) - (b - c) & \text{ak } \phi(\xi^\star, \boldsymbol{\theta}) < c \end{cases} \right) \right] = 0,$$

pričom $\mathcal{F}_{\phi(\cdot, \boldsymbol{\theta})}(\xi, \nu)$ označuje smerovú deriváciu kritéria $\phi(\cdot, \boldsymbol{\theta})$ v bode ξ a v smere ν. Výrazné zjednodušenie nastane ak je $\phi(\xi, \boldsymbol{\theta})$ spojitá náhodná premenná a kritérium $\bar{\phi}[M(\xi, \boldsymbol{\theta})] = \phi(\xi, \boldsymbol{\theta})$ je diferencovateľné na množine informačných matíc. Označme $G(\xi, \boldsymbol{\theta}) \in \mathbb{R}^{m \times m}$ gradient $\bar{\phi}$ vzhľadom na $M \in \mathbb{R}^{m \times m}$ v bode $M(\xi, \boldsymbol{\theta})$ a $k(\xi, \boldsymbol{\theta}) = \operatorname{tr}[M(\xi, \boldsymbol{\theta}) G(\xi, \boldsymbol{\theta})]$. Potom návrh $\xi^\star$ je CVaR-optimálny vtedy a len vtedy ak

$$0 = \max_{\substack{\mathbf{x} \in \mathcal{X} \\ b \in \mathbb{R}}} \left[b - \Phi_\alpha^Q(\xi^\star) + \right.$$
$$\left. + \frac{1}{\alpha} E \left(\begin{cases} 0 & \text{ak } \phi(\xi^\star, \boldsymbol{\theta}) > \Phi_\alpha^Q(\xi^\star) \\ \operatorname{tr}[M(\mathbf{x}, \boldsymbol{\theta}) G(\xi^\star, \boldsymbol{\theta})] - k(\xi^\star, \boldsymbol{\theta}) - (b - \Phi_\alpha^Q(\xi^\star)) & \text{ak } \phi(\xi^\star, \boldsymbol{\theta}) < \Phi_\alpha^Q(\xi^\star) \end{cases} \right) \right].$$

Ďalej sme sa v práci venovali metodike výpočtu optimálnych návrhov. Podobne ako v prácach Pflug (2000), Rockafellar a Uryasev (2000), Rockafellar a Uryasev (2002), Valenzuela et al. (2015) sme riešili úlohu nájdenia takého návrhu $\xi^\star$ a takého $c^\star$, že $(\xi^\star, c^\star) = \arg\max_{\xi \in \Xi, c \in \mathbb{R}} w_\alpha(\xi, c)$, kde $w_\alpha(\xi, c) = c + \frac{1}{\alpha} E[\min\{0, \phi(\xi, \boldsymbol{\theta}) - c\}]$. Na riešenie tejto úlohy sme sa rozhodli použiť Kelleyho metódu *cutting planes* z článku Kelley (1960). Najprv sme však museli vyjadriť subgradient funkcie w_α v bode $(\widetilde{\xi}, \widetilde{c})$, pričom sme dostali

$$\nabla w_\alpha\left(\widetilde{\xi}, \widetilde{c}\right) = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} + \frac{1}{\alpha} E \left[\begin{cases} \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix} & \text{ak } \phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) > \widetilde{c}, \\ \begin{pmatrix} \nabla_\xi \phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right) \\ -1 \end{pmatrix} & \text{inak} \end{cases} \right],$$

kde $\nabla_\xi \phi\left(\widetilde{\xi}, \boldsymbol{\theta}\right)$ je subgradient kritéria $\phi(\xi, \boldsymbol{\theta})$ vzhľadom na ξ v bode $\widetilde{\xi}$. Pokiaľ bolo zložité vyčísliť strednú hodnotu v subgradiente presne, použili sme jej aproximáciu na

základe Monte Carlo simulácií, pričom sme generovali nezávislé realizácie z apriórneho rozdelenia π (podobne postupovali aj Valenzuela et al. (2015) pri aproximácii strednej hodnoty v (5.4)) . Algoritmus, ktorý opäť riešil problém lineárneho programovania s nekonečne veľa ohraničeniami, sme potom otestovali na príkladoch, kde sme porovnali optimálne návrhy vzhľadom na priemerovacie, maximinné, kvantilové a CVaR krietérium.

Formulácia rozšírených kritériií optimality za účelom obmedzenia mylných odhadov v zovšeobecnených regresných modeloch

V článku Pázman a Pronzato (2014) a v Kap. 7 monografie Pronzato a Pázman (2013) sa autori venujú problému stability a jednoznačnosti odhadu metódou najmenších štvorcov v klasickej nelineárnej regresii. Tomuto problému môžeme predchádzať ešte na úrovni návrhu experimentu použitím ich rozšírených kritérií optimality, ktoré na jednej strane maximalizujú informáciu obsiahnutú v experimente, no na strane druhej minimalizujú možnosť mylného odhadu parametrov. Menovite autori zaviedli rozšírené kritériá E-, $\mathbf{c}$- a G-optimality, ktoré sa pri klasickom lineárnom regresnom modeli a v modeli s obmedzeným parametrickým priestorom správajú tak isto ako štandardné kritériá E-, $\mathbf{c}$- a G-optimality.

V práci sme rozšírili výsledky článku Pázman a Pronzato (2014) a Kap. 7 z Pronzato a Pázman (2013) pre účely stabilizovania odhadov metódou maximálnej vierohodnosti v zovšeobecnených regresných modeloch založených na exponenciálnej triede rozdelení. Kľúčovú úlohu tu zohráva I-divergencia, ktorá nesie informáciu o variabilite odhadu metódou maximálnej vierohodnosti ale súčasne odráža aj tzv. identifikovateľnosť parametra, čo je pojem často spomínaný v Kap. 7 z Pronzato a Pázman (2013). Niektoré výsledky sme publikovali v článku Burclová a Pázman (2016b).

Vo všeobecnosti sme v práci definovali (konkávne a pozitívne homogénne) rozšírené kritériá optimality pre nominálnu hodnotu parametra $\boldsymbol{\theta}^0$ nasledovne

$$\phi_\rho^{ext}\left(\xi,\boldsymbol{\theta}^0\right) \equiv \inf_{\left(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\in\Theta^k} \sum_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi\left(\mathbf{x}\right)\left[\frac{1}{\rho^2\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)}+K\right], \quad (5.6)$$

kde $k\in\mathbb{N}$ je dané číslo, $K\geq 0$ je ladiaca konštanta a $\rho:\mathbb{R}^{m\times(k+1)}\to\mathbb{R}$, $\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)\mapsto \rho\left(\boldsymbol{\theta}^0;\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k\right)$ je vzdialenosť medzi k-ticou bodov $\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k$ a nominálnou hodnotou $\boldsymbol{\theta}^0$ na parametrickom priestore Θ. Práve voľba vzdialenosti ρ, s čím súvisí aj voľba k, určuje, o rozšírenie akého kritéria sa jedná.

V práci sme sa intenzívnejšie venovali aj kritériám, ktoré sme nazvali pseudonormné. Pseudonorma $|||\cdot|||_P$ je zobrazenie, ktoré spĺňa všetky vlastnosti normy $\|\cdot\|$ okrem nasledovnej: $\|A\|=0\Leftrightarrow A=\mathbf{0}$. Často sme pritom pracovali s normou resp. pseudonormou

definovanou na priestore matíc $\mathbb{R}^{m\times k}$. Klasická (nerozšírená) verzia pseudonormných kritérií je

$$\phi_{\|\cdot\|_P}(\xi,\boldsymbol{\theta}) = \inf_{A\in\mathbb{R}^{m\times k}:\|A\|_P=1} \operatorname{tr}\left[A^\top M(\xi,\boldsymbol{\theta})\,A\right]. \tag{5.7}$$

Dette et al. (1995) sa venujú konvexnej verzii týchto kritérií s normou a nazývajú ich minimaxné kritériá (my sme použili názov pseudonormné kritériá na odlíšenie od maximinných kritérií, ktoré tu značia niečo iné). Článok Dette et al. (1995) okrem toho obsahuje dôležité tvrdenie o úlohe duálnej normy (viď napr. Bhatia (1997) pre definíciu duálnej normy) pri prechode medzi konvexnými a konkávnymi kritériami a ďalší relevantný výsledok sa týka faktu, že trieda L_p kritérií pre $p \geq 1$ patrí do triedy (5.7) a dostaneme ju použitím Schattenovej normy $\|\cdot\|_{\mathrm{S}(q)}$ (viď napr. Bhatia (1997) pre definíciu Schattenovej normy) pri $k = m$ a vhodne zvolenom q. Na základe Dette et al. (1995) sme teda mohli vyjadriť A-optimalitu a kritérium MV-optimality (viď napr. López-Fidalgo et al. (1998)) ako prvky triedy (5.7). Okrem toho sú prvkami triedy (5.7) aj ďalšie kritériá, o ktorých je to však dobre známe (viď napr. Kap. 5.1.2 z Pronzato a Pázman (2013)). V nasledujúcej tabuľke sme zhrnuli niektoré kritéria patriace do pseudonormnej triedy (5.7).

kritérium	k	pseudonorma
E	1	$\|\mathbf{u}\|_{\ell(2)} = \sqrt{\mathbf{u}^\top\mathbf{u}}$
MV	1	$\|\mathbf{u}\|_{\ell(\infty)} = \max_{i=1,\ldots,m}\|u_i\|$
$\mathbf{c}$	1	$\|\|\|\mathbf{u}\|\|\|_{P\mathbf{c}} = \left\|\mathbf{u}^\top\mathbf{c}\right\|$
G	1	$\|\|\|\mathbf{u}\|\|\|_{PG} = \max_{\mathbf{x}\in\mathcal{X}}\left\|\mathbf{u}^\top\mathbf{f}(\mathbf{x})\right\|$
A	m	$\|A\|_{\mathrm{S}(1)} = \operatorname{tr}[(A^\top A)^{1/2}]$
L	$m > k > 1$	$\|\|\|A\|\|\|_{PL} = \left\|\operatorname{tr}(A^\top L)\right\|$

Niektorí známi predstavitelia pseudonormnej triedy. Klasické definície jednotlivých kritérií možno nájsť napríklad v Kap. 5.1.2 z Pronzato a Pázman (2013).

Keď vzdialenosť $\rho(\cdot)$ z (5.6) definujeme pomocou pseudonormy, dostaneme rozšírené pseudonormné kritéria

$$\phi^{ext}_{\|\cdot\|_P}\left(\xi,\boldsymbol{\theta}^0\right) = \inf_{(\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^k)\in\Theta^k} \sum_{\mathbf{x}\in\mathcal{X}}\sum_{i=1}^{k} 2I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^i\right)\xi(\mathbf{x})\left[\frac{1}{\left\|\left(\boldsymbol{\theta}^0-\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^0-\boldsymbol{\theta}^k\right)\right\|_P^2}+K\right]. \tag{5.8}$$

Aplikovaním konkrétnych pseudonoriem z tabuľky vyššie dostávame rozšírenia známych kritérií E-, $\mathbf{c}$- a G-optimality, a tiež MV-, A- a L-optimality, ktoré v prvotnom článku Pázman a Pronzato (2014) neboli uvažované.

To, že kritériá v (5.8) sú naozaj rozšírením kritérií (5.7), vyplýva z nasledovných vlastností, ktoré boli pri splnení určitých podmienok v práci dokázané:

- v prípade klasického lineárneho regresného modelu s normálne rozdelenými chybami sa kritériá (5.8) zhodujú s tými v (5.7),
- ak parametrický priestor Θ je m-rozmerná guľa s polomerom r, tak kritériá v (5.8) konvergujú pre $r \to 0$ ku kritériám (5.7),
- ak model neobsahuje žiadne prekrytia v $\boldsymbol{\theta}^0$ pri danom návrhu ξ (teda $\sum_{\mathbf{x}\in\mathcal{X}} I_{\mathbf{x}}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}\right)\xi(\mathbf{x})$ konverguje k nule len pre $\boldsymbol{\theta} \to \boldsymbol{\theta}^0$), tak kritériá v (5.8) konvergujú pre $K \to \infty$ ku kritériám (5.7).

V dizertačnej práci sme rozobrali aj alternatívne všeobecnejšie definície rozšírených kritérií **c**- a G-optimality, ktoré sme rozpracovali v článku Burclová a Pázman (2016b) vychádzajúc z Pázman a Pronzato (2014).

Appendix

A Reformulation of L-optimality criterion

Theorem A1. *Let $L \neq \mathbf{0}$ be a matrix from $\mathbb{R}^{m\times k}$ and let $M \in \mathbb{R}^{m\times m}$ be a symmetric, positive semidefinite matrix. Then*

$$\inf_{Z\in\mathbb{R}^{m\times k}:\left|\mathrm{tr}\left(L^\top Z\right)\right|=1} \mathrm{tr}\left(Z^\top M Z\right) = \begin{cases} \frac{1}{\mathrm{tr}\left(L^\top M^- L\right)} & \text{if } \mathcal{C}(L) \subseteq \mathcal{C}(M) \\ 0 & \text{otherwise,} \end{cases}$$

where M^- denotes an arbitrary generalized inverse of M. Moreover, the infimum is always attained on $\left\{Z \in \mathbb{R}^{m\times k} : \left|\mathrm{tr}\left(L^\top Z\right)\right| = 1\right\}$.

Proof. The theorem is an extension of known property of **c**-optimality criterion, see e.g. Lemma 5.6 from Pronzato and Pázman (2013). In the proof we postponed similarly as Pronzato and Pázman (2013).

We start the proof with the case $\mathcal{C}(L) \subseteq \mathcal{C}(M)$, then $\mathrm{tr}\left(L^\top M^- L\right) = 0$ if and only if $L = \mathbf{0}$. Indeed, there is a matrix F such that $L = MF$ (see Harville, 2008, Lemma 4.2.2) and hence $\mathrm{tr}\left(L^\top M^- L\right) = 0 \Leftrightarrow \mathrm{tr}\left(F^\top M M^- M F\right) = 0 \Leftrightarrow \mathrm{tr}\left(F^\top M F\right) = 0 \Leftrightarrow \left\|M^{1/2}F\right\|_F^2 = 0 \Leftrightarrow M^{1/2}F = \mathbf{0} \Rightarrow L = \mathbf{0}$ and straightforwardly $L = \mathbf{0}$ implies $\mathrm{tr}\left(L^\top M^- L\right) = 0$. Moreover, one sees that the expression $L^\top M^- L$ is invariant to the choice of generalized inverse (see also Harville, 2008, Theorem 9.4.1).

We set $A^\top = L^\top M^+ M^{1/2}$, $B = M^{1/2}Z$, where M^+ is Moore-Penrose generalized inverse of M and $Z \in \mathbb{R}^{m\times k}$. According to Theorem 20.5.3(2) in Harville (2008) M^+ is symmetric matrix, and moreover, from Lemma 9.3.5 in Harville (2008), one has $L^\top = L^\top M^+ M$. The Cauchy-Schwarz inequality (Theorem 1.1) for matrices A and B gives $\mathrm{tr}^2\left(L^\top M^+ M Z\right) = \mathrm{tr}^2\left(L^\top Z\right) \leq \mathrm{tr}\left(L^\top M^+ L\right)\mathrm{tr}\left(Z^\top M Z\right)$ for any matrix $Z \in \mathbb{R}^{m\times k}$, which gives

$$\inf_{Z\in\mathbb{R}^{m\times k}:\mathrm{tr}\left(L^\top Z\right)\neq 0} \frac{\mathrm{tr}\left(Z^\top M Z\right)}{\mathrm{tr}^2\left(L^\top Z\right)} \geq \frac{1}{\mathrm{tr}\left(L^\top M^+ L\right)} = \frac{1}{\mathrm{tr}\left(L^\top M^- L\right)}, \tag{A.1}$$

since the last expression does not depend on the choice of generalized inverse. Now set

$Z^* = M^+ L$. It follows that $\mathrm{tr}\left(L^\top Z^*\right) = \mathrm{tr}\left(L^\top M^+ L\right) \neq 0$. We obtain

$$\frac{\mathrm{tr}\left(Z^{*\top} M Z^*\right)}{\mathrm{tr}^2\left(L^\top Z^*\right)} = \frac{\mathrm{tr}\left(L^\top M^+ M M^+ L\right)}{\mathrm{tr}^2\left(L^\top M^+ L\right)} = \frac{1}{\mathrm{tr}\left(L^\top M^+ L\right)} = \frac{1}{\mathrm{tr}\left(L^\top M^- L\right)}. \tag{A.2}$$

It follows from (A.1) and (A.2) that

$$\inf_{Z \in \mathbb{R}^{m\times k}: \mathrm{tr}\left(L^\top Z\right) \neq 0} \frac{\mathrm{tr}\left(Z^\top M Z\right)}{\mathrm{tr}^2\left(L^\top Z\right)} = \inf_{Z \in \mathbb{R}^{m\times k}: \left|\mathrm{tr}\left(L^\top Z\right)\right| = 1} \mathrm{tr}\left(Z^\top M Z\right) = \frac{1}{\mathrm{tr}\left(L^\top M^- L\right)}.$$

Now suppose that $\mathcal{C}(L) \nsubseteq \mathcal{C}(M)$. Then, according to Lemma 4.2.1 and Theorem 12.1.3 in Harville (2008), there are two matrices $L_1 \neq \mathbf{0}$ and L_2 such that $L = L_1 + L_2$, $\mathcal{C}(L_2) \subseteq \mathcal{C}(M)$ and $\mathrm{tr}\left(L_1^\top G\right) = 0$ for any matrix $G \in \mathbb{R}^{m\times k}$ such that $\mathcal{C}(G) \subseteq \mathcal{C}(M)$. One sees that $\mathrm{tr}\left(L^\top L_1\right) = \mathrm{tr}\left(L_1^\top L_1 + L_2^\top L_1\right) = \mathrm{tr}\left(L_1^\top L_1\right) = \|L_1\|_F^2 > 0$ since $L_1 \neq \mathbf{0}$. We set $Z^{**} = \frac{L_1}{\mathrm{tr}\left(L^\top L_1\right)}$. Then $\mathrm{tr}\left(L^\top Z^{**}\right) = 1$ and

$$0 \leq \inf_{Z \in \mathbb{R}^{m\times k}: \left|\mathrm{tr}\left(L^\top Z\right)\right| = 1} \mathrm{tr}\left(Z^\top M Z\right) \leq \left(Z^{**\top} M Z^{**}\right) = \frac{\mathrm{tr}\left(L_1^\top M L_1\right)}{\mathrm{tr}^2\left(L^\top L_1\right)} = 0.$$

□

B Some examples of distributions from the exponential family

Here we provide an overview of some well known distributions from the exponential family and we derive the corresponding Fisher information matrices and I-divergences. Where possible, we check our results with Atkinson et al. (2014) (Fisher information matrices in their Tables 2 and 4, and in Sect. 4.3) and Vajda and van der Meulen (1998) (I-divergences). Here we will always use the notation $\boldsymbol{\vartheta}$ for the usual parametrization in the given exponential family.

B.1 Binomial distribution $Bin(n,p)$

$n \in \mathbb{N}$ is given, $p \in (0,1)$, $y \in \{0,1,\ldots,n\}$

$$\begin{aligned} f(y,p) &= \binom{n}{y} p^y (1-p)^{n-y} \\ &= \exp\left\{ \ln\binom{n}{y} + y \ln\left(\frac{p}{1-p}\right) + n \ln(1-p) \right\} \end{aligned}$$

- $\vartheta = p$
- $t(y) = y$
- $\gamma(p) = \ln\left(\frac{p}{1-p}\right)$
- $\psi(y) = -\ln\binom{n}{y}$
- $\kappa(\gamma) = n \ln(1+e^{\gamma})$
- $\overline{\mu}(\gamma) = n\frac{e^{\gamma}}{1+e^{\gamma}}$
- $\overline{\mu}[\gamma(p)] = np$
- $M_{\boldsymbol{\vartheta}} = \frac{n}{p(1-p)}$
- $I\left[\gamma\left(p^0\right), \gamma(p)\right] = n\left[p^0 \ln\left(\frac{p^0}{p}\right) + \left(1-p^0\right) \ln\left(\frac{1-p^0}{1-p}\right)\right]$
- $I\left(\gamma^0, \gamma\right) = n\left[\left(\gamma^0 - \gamma\right)\frac{e^{\gamma^0}}{1+e^{\gamma^0}} + \ln\frac{1+e^{\gamma}}{1+e^{\gamma^0}}\right]$

B.2 Poisson distribution $Po(\lambda)$

$\lambda > 0$, $y \in \mathbb{N} \cup \{0\}$

$$\begin{aligned} f(y,\lambda) &= e^{-\lambda}\frac{\lambda^y}{y!} \\ &= \exp\{-\ln(y!) + y\ln(\lambda) - \lambda\} \end{aligned}$$

- $\vartheta = \lambda$
- $t(y) = y$
- $\gamma(\lambda) = \ln(\lambda)$
- $\psi(y) = \ln(y!)$
- $\kappa(\gamma) = e^{\gamma}$
- $\overline{\mu}(\gamma) = e^{\gamma}$
- $\overline{\mu}[\gamma(\lambda)] = \lambda$
- $M_{\boldsymbol{\vartheta}} = 1/\lambda$
- $I\left[\gamma\left(\lambda^0\right), \gamma(\lambda)\right] = \lambda^0\left[\frac{\lambda}{\lambda^0} - 1 - \ln\left(\frac{\lambda}{\lambda^0}\right)\right]$
- $I\left(\gamma^0, \gamma\right) = \left(\gamma^0 - \gamma\right)e^{\gamma^0} + e^{\gamma} - e^{\gamma^0}$

B.3 Geometric distribution $Ge(p)$

$p \in (0,1)$, $y \in \mathbb{N} \cup \{0\}$

$$\begin{aligned} f(y,p) &= (1-p)p^y \\ &= \exp\{y\ln(p) + \ln(1-p)\} \end{aligned}$$

- $\vartheta = p$
- $t(y) = y$
- $\gamma(p) = \ln(p)$
- $\psi(y) = 0$
- $\kappa(\gamma) = -\ln(1 - e^{\gamma})$
- $\overline{\mu}(\gamma) = \frac{e^{\gamma}}{1-e^{\gamma}}$
- $\overline{\mu}[\gamma(p)] = \frac{p}{1-p}$
- $M_{\boldsymbol{\vartheta}} = \frac{1}{p(1-p)^2}$
- $I\left[\gamma\left(p^0\right), \gamma(p)\right] = \frac{1}{1-p^0}\left[p^0\ln\left(\frac{p^0}{p}\right) + \left(1-p^0\right)\ln\left(\frac{1-p^0}{1-p}\right)\right]$
- $I\left(\gamma^0, \gamma\right) = \left(\gamma^0 - \gamma\right)\frac{e^{\gamma^0}}{1-e^{\gamma^0}} + \ln\left(\frac{1-e^{\gamma^0}}{1-e^{\gamma}}\right)$

B.4 Negative binomial distribution $NBin(r,p)$

$p \in (0,1)$, $r \in \mathbb{N}$ is given, $y \in \mathbb{N} \cup \{0\}$

$$\begin{aligned} f(y,p) &= \binom{y+r-1}{y} p^y (1-p)^r \\ &= \exp\left\{ \ln \binom{y+r-1}{y} + y \ln(p) + r \ln(1-p) \right\} \end{aligned}$$

- $\vartheta = p$
- $t(y) = y$
- $\gamma(p) = \ln(p)$
- $\psi(y) = -\ln\binom{y+r-1}{y}$
- $\kappa(\gamma) = -r \ln(1-e^{\gamma})$
- $\overline{\mu}(\gamma) = \frac{re^{\gamma}}{1-e^{\gamma}}$
- $\overline{\mu}[\gamma(p)] = \frac{rp}{1-p}$
- $M_{\boldsymbol{\vartheta}} = \frac{r}{p(1-p)^2}$
- $I\left[\gamma\left(p^0\right), \gamma(p)\right] = \frac{r}{1-p^0}\left[p^0 \ln\left(\frac{p^0}{p}\right) + \left(1-p^0\right) \ln\left(\frac{1-p^0}{1-p}\right)\right]$
- $I\left(\gamma^0, \gamma\right) = r\left[\left(\gamma^0 - \gamma\right)\frac{e^{\gamma^0}}{1-e^{\gamma^0}} + \ln\left(\frac{1-e^{\gamma^0}}{1-e^{\gamma}}\right)\right]$

B.5 Exponential distribution $Exp(\lambda)$

$\lambda > 0$, $y \geq 0$

$$\begin{aligned} f(y,\lambda) &= \lambda e^{-\lambda y} \\ &= \exp\{-\lambda y + \ln \lambda\} \end{aligned}$$

- $\vartheta = \lambda$
- $t(y) = -y$
- $\gamma(\lambda) = \lambda$
- $\psi(y) = 0$
- $\kappa(\gamma) = -\ln(\gamma)$
- $\overline{\mu}(\gamma) = -1/\gamma$
- $\overline{\mu}[\gamma(\lambda)] = -1/\lambda$
- $M_{\boldsymbol{\vartheta}} = \frac{1}{\lambda^2}$
- $I\left[\gamma\left(\lambda^0\right), \gamma(\lambda)\right] = -1 + \frac{\lambda}{\lambda^0} - \ln\left(\frac{\lambda}{\lambda^0}\right)$
- $I\left(\gamma^0, \gamma\right) = -1 + \frac{\gamma}{\gamma^0} - \ln\left(\frac{\gamma}{\gamma^0}\right)$

B.6 Weibull distribution $W(\lambda, k)$

$\lambda > 0$, $k > 0$ is given, $y \geq 0$

$$\begin{aligned} f(y,\lambda) &= \frac{k}{\lambda}\left(\frac{y}{\lambda}\right)^{k-1} e^{-(y/\lambda)^k} \\ &= \exp\left\{\ln(k) + (k-1)\ln(y) - y^k\lambda^{-k} - k\ln\lambda\right\} \end{aligned}$$

- $\vartheta = \lambda$
- $t(y) = -y^k$
- $\gamma(\lambda) = \lambda^{-k}$
- $\psi(y) = -\ln(k) - (k-1)\ln(y)$
- $\kappa(\gamma) = -\ln(\gamma)$
- $\overline{\mu}(\gamma) = -1/\gamma$
- $\overline{\mu}[\gamma(\lambda)] = -\lambda^k$
- $M_{\boldsymbol{\vartheta}} = \frac{k^2}{\lambda^2}$
- $I\left[\gamma\left(\lambda^0\right), \gamma(\lambda)\right] = -1 + \left(\frac{\lambda^0}{\lambda}\right)^k - k\ln\left(\frac{\lambda^0}{\lambda}\right)$
- $I\left(\gamma^0, \gamma\right) = -1 + \frac{\gamma}{\gamma^0} - \ln\left(\frac{\gamma}{\gamma^0}\right)$

B.7 Pareto distribution $Pa(\alpha, \sigma)$

$\alpha > 0$, $\sigma > 0$ is given, $y \geq \sigma$

$$\begin{aligned} f(y,\alpha) &= \frac{\alpha}{y}\left(\frac{y}{\sigma}\right)^{-\alpha} \\ &= \exp\left\{-(\alpha+1)\ln(y) + \ln\alpha + \alpha\ln\sigma\right\} \end{aligned}$$

- $\vartheta = \alpha$
- $t(y) = -\ln(y)$
- $\gamma(\alpha) = \alpha + 1$
- $\psi(y) = 0$
- $\kappa(\gamma) = -\ln(\gamma - 1) - (\gamma - 1)\ln\sigma$
- $\overline{\mu}(\gamma) = -1/(\gamma - 1) - \ln\sigma$
- $\overline{\mu}[\gamma(\alpha)] = -1/\alpha - \ln\sigma$
- $M_{\boldsymbol{\vartheta}} = 1/\alpha^2$
- $I\left[\gamma\left(\alpha^0\right), \gamma(\alpha)\right] = \ln\frac{\alpha^0}{\alpha} + \left(-\frac{1}{\alpha^0} - \ln\sigma\right)\left(\alpha^0 - \alpha\right) + \left(\alpha^0 - \alpha\right)\ln\sigma$
- $I\left(\gamma^0, \gamma\right) = \ln\frac{\gamma^0 - 1}{\gamma - 1} + \left(-\frac{1}{\gamma^0 - 1} - \ln\sigma\right)\left(\gamma^0 - \gamma\right) + \left(\gamma^0 - \gamma\right)\ln\sigma$

B.8 Normal distribution $\mathcal{N}\left(\nu,\sigma^2\right)$ with σ^2 given

$\nu \in \mathbb{R}$, $\sigma^2 > 0$ is given, $y \in \mathbb{R}$

$$\begin{aligned} f(y,\nu) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(y-\nu)^2\right\} \\ &= \exp\left\{-\ln\left(\sqrt{2\pi}\sigma\right) - \frac{y^2}{2\sigma^2} + \frac{y\nu}{\sigma^2} - \frac{\nu^2}{2\sigma^2}\right\} \end{aligned}$$

- $\vartheta = \nu$
- $t(y) = y$
- $\gamma(\nu) = \frac{\nu}{\sigma^2}$
- $\psi(y) = \ln\left(\sqrt{2\pi}\sigma\right) + \frac{y^2}{2\sigma^2}$
- $\kappa(\gamma) = \frac{\sigma^2\gamma^2}{2}$
- $\overline{\mu}(\gamma) = \sigma^2\gamma$
- $\overline{\mu}[\gamma(\nu)] = \nu$
- $M_{\boldsymbol{\vartheta}} = \frac{1}{\sigma^2}$
- $I\left[\gamma\left(\nu^0\right),\gamma(\nu)\right] = \frac{1}{2\sigma^2}\left(\nu^0-\nu\right)^2$
- $I\left(\gamma^0,\gamma\right) = \frac{\sigma^2}{2}\left(\gamma^0-\gamma\right)^2$

B.9 Normal distribution $\mathcal{N}\left(\nu,\sigma^2\right)$

$\nu \in \mathbb{R}$, $\sigma^2 > 0$, $y \in \mathbb{R}$

$$\begin{aligned} f\left(y,\nu,\sigma^2\right) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(y-\nu)^2\right\} \\ &= \exp\left\{-\ln\left(\sqrt{2\pi}\right) + \left(y,-y^2\right)\begin{pmatrix}\nu/\sigma^2\\ 1/2\sigma^2\end{pmatrix} - \ln\sigma - \frac{\nu^2}{2\sigma^2}\right\} \end{aligned}$$

- $\boldsymbol{\vartheta} = \begin{pmatrix}\nu\\ \sigma^2\end{pmatrix}$
- $t(y) = \begin{pmatrix}y\\ -y^2\end{pmatrix}$
- $\gamma\left[\left(\nu,\sigma^2\right)^\top\right] = \begin{pmatrix}\nu/\sigma^2\\ 1/\left(2\sigma^2\right)\end{pmatrix}$
- $\psi(y) = \ln\left(\sqrt{2\pi}\right)$
- $\kappa(\boldsymbol{\gamma}) = -\frac{1}{2}\ln(2\gamma_2) + \frac{\gamma_1^2}{4\gamma_2}$
- $\overline{\mu}(\boldsymbol{\gamma}) = \begin{pmatrix}\frac{\gamma_1}{2\gamma_2}\\ -\frac{1}{2\gamma_2} - \frac{\gamma_1^2}{4\gamma_2^2}\end{pmatrix}$
- $\overline{\mu}\left\{\gamma\left[\left(\nu,\sigma^2\right)^\top\right]\right\} = \begin{pmatrix}\nu\\ -\sigma^2-\nu^2\end{pmatrix}$
- $M_{\boldsymbol{\vartheta}} = \frac{1}{\sigma^2}\begin{pmatrix}1 & 0\\ 0 & 1/\left(2\sigma^2\right)\end{pmatrix}$
- $I\left\{\gamma\left[\left(\nu^0,\left(\sigma^0\right)^2\right)^\top\right],\gamma\left[\left(\nu,\sigma^2\right)^\top\right]\right\} = \frac{1}{2}\left[\frac{\left(\nu-\nu^0\right)^2}{\sigma^2} + \frac{\left(\sigma^0\right)^2}{\sigma^2} - 1 - \ln\left(\frac{\left(\sigma^0\right)^2}{\sigma^2}\right)\right]$
- $I\left(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}\right) = \frac{\gamma_1^0\left(\gamma_1^0-\gamma_1\right)}{2\gamma_2^0} + \left(\gamma_2^0-\gamma_2\right)\left[-\left(\frac{\gamma_1^0}{2\gamma_2^0}\right)^2 - \frac{1}{2\gamma_2^0}\right] + \frac{1}{2}\ln\left(\frac{\gamma_2^0}{\gamma_2}\right) + \frac{(\gamma_1)^2}{4\gamma_2} - \frac{\left(\gamma_1^0\right)^2}{4\gamma_2^0}$

B.10 Gamma distribution $G(k,\alpha)$

$k>0$, $\alpha>0$, $y>0$

$$f(y,k,\alpha)=\frac{y^{k-1}e^{-\alpha y}\alpha^{k}}{\Gamma(k)}$$
$$=\exp\left\{(\ln y,-y)\begin{pmatrix}k-1\\ \alpha\end{pmatrix}+k\ln(\alpha)-\ln\left[\Gamma(k)\right]\right\}$$

- $\boldsymbol{\vartheta}=\begin{pmatrix}k\\ \alpha\end{pmatrix}$
- $t(y)=\begin{pmatrix}\ln y\\ -y\end{pmatrix}$
- $\gamma\left[(k,\alpha)^{\top}\right]=\begin{pmatrix}k-1\\ \alpha\end{pmatrix}$
- $\psi(y)=0$
- $\kappa(\boldsymbol{\gamma})=-(\gamma_1+1)\ln(\gamma_2)+\ln\left[\Gamma(\gamma_1+1)\right]$
- $\overline{\mu}(\boldsymbol{\gamma})=\begin{pmatrix}-\ln(\gamma_2)+\Psi(\gamma_1+1)\\ -\frac{1+\gamma_1}{\gamma_2}\end{pmatrix}$
- $\overline{\mu}\left\{\gamma\left[(k,\alpha)^{\top}\right]\right\}=\begin{pmatrix}-\ln\alpha+\Psi(k)\\ -\frac{k}{\alpha}\end{pmatrix}$
- $M_{\boldsymbol{\vartheta}}=\begin{pmatrix}\Psi'(k) & -\frac{1}{\alpha}\\ -\frac{1}{\alpha} & \frac{k}{\alpha^2}\end{pmatrix}$
- $I\left\{\gamma\left[\left(k^0,\alpha^0\right)^{\top}\right],\gamma\left[(k,\alpha)^{\top}\right]\right\}=\left(k^0-k\right)\Psi\left(k^0\right)+k\ln\left(\frac{\alpha_0}{\alpha}\right)+\ln\left[\frac{\Gamma(k)}{\Gamma(k^0)}\right]+k_0\left(\frac{\alpha}{\alpha^0}-1\right)$
- $I\left(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}\right)=\left(\gamma_1^0-\gamma_1\right)\Psi\left(\gamma_1^0+1\right)+(\gamma_1+1)\ln\left(\frac{\gamma_2^0}{\gamma_2}\right)+\ln\left[\frac{\Gamma(\gamma_1+1)}{\Gamma(\gamma_1^0+1)}\right]+\left(1+\gamma_1^0\right)\left(\frac{\gamma_2}{\gamma_2^0}-1\right)$

$\Gamma(k)=\int_0^\infty z^{k-1}e^{-z}dz$ is the gamma function, $\Psi(k)=\frac{\partial\ln\Gamma(k)}{\partial k}$ is the digamma function and $\Psi'(k)=\frac{\partial\Psi(k)}{\partial k}=\frac{\partial^2\ln\Gamma(k)}{\partial k^2}$ is the trigamma function, see Abramowitz and Stegun (1972) Sects. 6.3.1 and 6.4.1.

B.11 Beta distribution $Be(a,b)$

$a>0$, $b>0$, $y\in(0,1)$

$$f(y,a,b)=\frac{1}{B(a,b)}y^{a-1}(1-y)^{b-1}$$
$$=\exp\left\{(\ln(y),\ln(1-y))\begin{pmatrix}a-1\\b-1\end{pmatrix}-\ln B(a,b)\right\}$$

- $\boldsymbol{\vartheta}=\begin{pmatrix}a\\b\end{pmatrix}$
- $t(y)=\begin{pmatrix}\ln(y)\\\ln(1-y)\end{pmatrix}$
- $\gamma\left[(a,b)^{\top}\right]=\begin{pmatrix}a-1\\b-1\end{pmatrix}$
- $\psi(y)=0$
- $\kappa(\boldsymbol{\gamma})=\ln\left[B(\gamma_1+1,\gamma_2+1)\right]$
- $\overline{\mu}(\boldsymbol{\gamma})=$
  $\begin{pmatrix}\Psi(\gamma_1+1)-\Psi(\gamma_1+\gamma_2+2)\\\Psi(\gamma_2+1)-\Psi(\gamma_1+\gamma_2+2)\end{pmatrix}$
- $\overline{\mu}\left\{\gamma\left[(a,b)^{\top}\right]\right\}=$
  $\begin{pmatrix}\Psi(a)-\Psi(a+b)\\\Psi(b)-\Psi(a+b)\end{pmatrix}$
- $M_{\boldsymbol{\vartheta}}=$
  $\begin{pmatrix}\Psi'(a)-\Psi'(a+b) & -\Psi'(a+b)\\-\Psi'(a+b) & \Psi'(b)-\Psi'(a+b)\end{pmatrix}$
- $I\left\{\gamma\left[\left(a^0,b^0\right)^{\top}\right],\gamma\left[(a,b)^{\top}\right]\right\}=$
  $\ln\left[\frac{B(a,b)}{B(a^0,b^0)}\right]+\Psi\left(a^0\right)\left(a^0-a\right)+$
  $\Psi\left(b^0\right)\left(b^0-b\right)-$
  $\Psi\left(a^0+b^0\right)\left(a^0-a+b^0-b\right)$
- $I\left(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}\right)=\ln\left[\frac{B(\gamma_1+1,\gamma_2+1)}{B\left(\gamma_1^0+1,\gamma_2^0+1\right)}\right]+$
  $\Psi\left(\gamma_1^0+1\right)\left(\gamma_1^0-\gamma_1\right)+$
  $\Psi\left(\gamma_2^0\right)\left(\gamma_2^0-\gamma_2\right)-$
  $\Psi\left(\gamma_1^0+\gamma_2^0+2\right)\left(\gamma_1^0-\gamma_1+\gamma_2^0-\gamma_2\right)$

$B(a,b)=\int_0^1 z^{a-1}(1-z)^{b-1}dz$ is the beta function, $\Psi(k)=\frac{\partial\ln\Gamma(k)}{\partial k}$ is the digamma function and $\Psi'(k)=\frac{\partial\Psi(k)}{\partial k}=\frac{\partial^2\ln\Gamma(k)}{\partial k^2}$ is the trigamma function, see Abramowitz and Stegun (1972) Sects. 6.3.1 and 6.4.1.

B.12 Multinomial distribution $Mult(n, p_1, \ldots, p_{l-1})$

$n \in \mathbb{N}$ is given, $p_i > 0$ for $i = 1, \ldots, l-1$ and $\sum_{i=1}^{l-1} p_i < 1$, $y_i \in \mathbb{N} \cup \{0\}$ for $i = 1, \ldots, l-1$ and $\sum_{i=1}^{l-1} y_i \leq n$

$$f(y_1, \ldots, y_{l-1}, p_1, \ldots, p_{l-1}) = \frac{n!}{y_1! \ldots y_{l-1}! y_l!} p_1^{y_1} \ldots p_{l-1}^{y_{l-1}} p_l^{y_l}$$

$$= \exp\left\{ \ln(n!) - \ln[y_1! \ldots y_{l-1}! y_l!] + (y_1, \ldots, y_{l-1}) \begin{pmatrix} \ln\left(\frac{p_1}{p_l}\right) \\ \vdots \\ \ln\left(\frac{p_{l-1}}{p_l}\right) \end{pmatrix} + n \ln(p_l) \right\},$$

where $p_l = 1 - \sum_{i=1}^{l-1} p_i$ and $y_l = n - \sum_{i=1}^{l-1} y_i$.

- $\boldsymbol{\vartheta} = \begin{pmatrix} p_1 \\ \vdots \\ p_{l-1} \end{pmatrix}$
- $t(y) = \begin{pmatrix} y_1 \\ \vdots \\ y_{l-1} \end{pmatrix}$
- $\gamma\left[(p_1, \ldots, p_{l-1})^\top\right] = \begin{pmatrix} \ln\left(\frac{p_1}{p_l}\right) \\ \vdots \\ \ln\left(\frac{p_{l-1}}{p_l}\right) \end{pmatrix}$
- $\psi(y) = -\ln n! + \ln(y_1! \ldots y_{l-1}! y_l!)$
- $\kappa(\boldsymbol{\gamma}) = -n \ln\left(\frac{1}{1+\sum_{i=1}^{l-1} e^{\gamma_i}}\right)$
- $\overline{\mu}(\boldsymbol{\gamma}) = n \begin{pmatrix} \frac{e^{\gamma_1}}{1+\sum_{i=1}^{l-1} e^{\gamma_i}} \\ \vdots \\ \frac{e^{\gamma_{l-1}}}{1+\sum_{i=1}^{l-1} e^{\gamma_i}} \end{pmatrix}$
- $\overline{\mu}\left\{\gamma\left[(p_1, \ldots, p_{l-1})^\top\right]\right\} = n \begin{pmatrix} p_1 \\ \vdots \\ p_{l-1} \end{pmatrix}$
- $M_{\boldsymbol{\vartheta}} = \frac{n}{p_l} \begin{pmatrix} \frac{p_1+p_l}{p_1} & 1 & \ldots & 1 \\ 1 & \frac{p_2+p_l}{p_2} & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & \frac{p_{l-1}+p_l}{p_{l-1}} \end{pmatrix}$
- $I\left\{\gamma\left[\left(p_1^0, \ldots, p_{l-1}^0\right)^\top\right], \gamma\left[(p_1, \ldots, p_{l-1})^\top\right]\right\}$
  $= n\left\{\sum_{i=1}^{l-1}\left[p_i^0 \ln\left(\frac{p_i^0}{p_i}\right)\right] + p_l^0 \ln\left(\frac{p_l^0}{p_l}\right)\right\}$
- $I\left(\boldsymbol{\gamma}^0, \boldsymbol{\gamma}\right) =$
  $n\left\{\sum_{i=1}^{l-1}\left[\frac{(\gamma_i^0 - \gamma_i) e^{\gamma_i^0}}{1+\sum_{j=1}^{l-1} e^{\gamma_j^0}}\right] + \ln\left[\frac{1+\sum_{j=1}^{l-1} e^{\gamma_j}}{1+\sum_{k=1}^{l-1} e^{\gamma_k^0}}\right]\right\}$

Bibliography

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Applied Mathematics Series, Vol. 55. National Bureau of Standards. tenth printing.

Atkinson, A. C., Chaloner, K., Herzberg, A. M., and Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics*, 49:325–337.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Atkinson, A. C., Fedorov, V. V., Herzberg, A. M., and Zhang, R. (2014). Elemental information matrices and optimal experimental design for generalized regression models. *J. Stat. Plan. Inference*, 144:81–91.

Atkinson, A. C. and Woods, D. C. (2015). Designs for generalized linear models. In Dean, A., Morris, M., Stufken, J., and Bingham, D., editors, *Handbook of Design and Analysis of Experiments*, chapter 13. Chapman and Hall/CRC Press.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley.

Berkelaar, M., Eikland, K., and Notebaert, P. (2004). lp_solve: Open source (mixed-integer) linear programming system. http://lpsolve.sourceforge.net/5.5/index.htm. Version 5.1.0.0 dated 1 May 2004.

Bhatia, R. (1997). *Matrix Analysis*. Graduate Texts in Mathematics, Vol. 169. Springer Science + Business Media, New York.

Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 3rd edition.

Bonnans, J. F., Gilbert, J. C., Lemaréchal, C., and Sagastizábal, C. A. (2006). *Numerical Optimization. Theoretical and Practical Aspects*. Springer, Heidelberg, 2nd edition.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. IMS Lecture Notes Monogr. Ser., Volume 9. Institute of Mathematical Statistics, Hayward, CA.

Burclová, K. and Pázman, A. (2016a). Optimal design of experiments via linear programming. *Stat. Papers*, 57:893–910.

Burclová, K. and Pázman, A. (2016b). Optimum design via I-divergence for stable estimation in generalized regression models. In Kunert, J., Müller, C. H., and Atkinson, A. C., editors, *mODa 11—Advances in Model-Oriented Design and Analysis*, pages 55–62. Springer.

Dette, H., Heiligers, B., and Studden, W. J. (1995). Minimax designs in linear regression models. *The Annals of Statistics*, 23(1):30–40.

Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*. Chapman & Hall/ CRC, London.

Efron, B. (1978). The geometry of exponential families. *Ann. Stat.*, 6(2):362–376.

Fedorov, V. V. and Leonov, S. L. (2014). *Optimal Design for Nonlinear Response Models*. CRC Press.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Volume II.* Wiley, New York, 2nd edition.

Filová, L. and Harman, R. (2013). Criterion-robust experimental designs for the quadratic regression on a square and a cube. *Commun Stat Theory Methods*, 42(11):2044–2055.

Filová, L., Trnovská, M., and Harman, R. (2012). Computing maximin efficient experimental designs using the methods of semidefinite programming. *Metrika*, 75(5):709–719.

Gentle, J. E. (2007). *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer Science + Business Media, New York.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3rd edition.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. R. Statist. Soc. B*, 46(2):149–192.

Gribik, P. R. and Kortanek, K. O. (1977). Equivalence theorems and cutting plane algorithms for a class of experimental design problems. *SIAM J. Appl. Math*, 32(1):232–259.

Guerra, J. (2016). *Optimisation multi-objectif sous incertitudes de phénomènes de thermique transitoire*. PhD thesis, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE).

Harman, R. (2004). Minimal efficiency of designs under the class of orthogonally invariant information criteria. *Metrika*, 60:137–153.

Harman, R. (2008). Equivalence theorem for Schur optimality of experimental designs. *Journal of Statistical Planning and Inference*, 138:1201–1209.

Harman, R. and Jurík, T. (2008). Computing c-optimal experimental designs using the simplex method of linear programming. *Comput Stat Data Anal*, 53:247–254.

Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*. Springer Science + Business Media, New York.

Hosseini, M. (2009). *Statistical models for agroclimate risk analysis*. PhD thesis, The University of British Columbia.

Hosseini, R. (2010). "Quantiles symmetry". *arXiv preprint*. Preprint Available at arXiv:1004.0540.

Jørgensen, B. (1984). The delta algorithm and GLIM. *International Statistical Review*, 52:283–300.

Kelley, J. E. (1960). The cutting-plane method for solving convex programs. *J Soc Indust Appl Math*, 8:703–712.

Khuri, A. I., Mukherjee, B., Sinha, B. K., and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science*, 21(3):376–399.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann Stat*, 2(5):849–879.

Kiefer, J. (1975). Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, 62(2):277–288.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *Ann Math Stat*, 30(2):271–294.

Kullback, S. (1997). *Information Theory and Statistics*. Dover Publications, Inc., Mineola, N.Y.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann Math Stat*, 22(1):79–86.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag, New York, 2nd edition.

López-Fidalgo, J., Tommasi, C., and Trandafir, P. C. (2007). An optimal experimental design criterion for discriminating between non-normal models. *J. R. Stat. Soc. B*, 69(2):231–242.

López-Fidalgo, J., Torsney, B., and Ardanuy, R. (1998). MV-optimization in weighted linear regression. In Atkinson, A. C., Pronzato, L., and Winn, H. P., editors, *MODA 5—Advances in Model-Oriented Data Analysis and Experimental Design*, pages 39–50. Physica, Heidelberg.

lp_solve and Konis., K. (2016). *lpSolveAPI: R Interface to 'lp_solve' Version 5.5.2.0*. R package version 5.5.2.0-17. https://CRAN.R-project.org/package=lpSolveAPI.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability, Vol. 37. Chapman & Hall/CRC, 2nd edition.

Morrow, J. (2013). Norms on $\mathbb{R}^n$. https://sites.math.washington.edu/~morrow/334_13/norms2.pdf. Accessed: 2017-12-11.

Nelder, J. A. (1975). Announcement by the working party on statistical computing: GLIM (Generalized Linear Interactive Modelling Program). *J. R. Statist. Soc. C (Applied Statistics)*, 24(2):259–261.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc. A*, 135(3):370–384.

Nesterov, Y. (2004). *Introductory Lectures on Convex Optimization: A Basic Course*. Springer Science + Business Media, New York.

Pázman, A. (1986). *Foundations of Optimum Experimental Design*. D. Reidel, Dordrecht.

Pázman, A. (1993). *Nonlinear Statistical Models*. Kluwer Academic Publishers, Dordrecht.

Pázman, A. (2015). 3-MAM-024 Nonlinear statistical models. (personal communication). Faculty of Mathematics, Physics and Informatics. Comenius University.

Pázman, A. and Lacko, V. (2012). *Prednášky z regresných modelov*. Univerzita Komenského v Bratislave. (In slovak language).

Pázman, A. and Pronzato, L. (2007). Quantile and probability-level criteria for nonlinear experimental design. In López-Fidalgo, J., Rodríguez-Díaz, J. M., and Torsney, B., editors, *mODa 8—Advances in Model-Oriented Design and Analysis*, pages 157–164. Physica-Verlag, Heidelberg.

Pázman, A. and Pronzato, L. (2014). Optimum design accounting for the global nonlinear behavior of the model. *Ann Stat*, 42(4):1426–1451.

Pflug, G. C. (2000). Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. In Uryasev, S. P., editor, *Probabilistic Constrained Optimization. Nonconvex Optimization and Its Applications*, volume 49, pages 272–281. Springer, Boston, MA. Available at http://www.pacca.info/public/files/docs/public/finance/Active%20Risk%20Management/gpfl.pdf. Accessed: 2018-04-11.

Pronzato, L. and Pázman, A. (2013). *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Lecture Notes in Statistics, Vol. 212. Springer, New York, Heidelberg.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41. Available at http://www.ise.ufl.edu/uryasev/files/2011/11/CVaR1_JOR.pdf. Accessed: 2018-04-11.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26:1443–1471.

Shimizu, K. and Aiyoshi, E. (1980). Necessary conditions for min-max problems and algorithms by a relaxation procedure. *IEEE Transactions on Automatic Control*, 25(1):62–66.

Steele, J. M. (2004). *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press, New York.

Trench, W. F. (2003). *Introduction to Real Analysis*. Pearson Education. (free online version 2013 http://digitalcommons.trinity.edu/mono/7).

Vajda, I. and van der Meulen, E. (1998). Global statistical information in exponential experiments and selection of exponential models. *Applications of Mathematics*, 43(1):23–51.

Valenzuela, P. E., Rojas, C. R., and Hjalmarsson, H. (2015). Uncertainty in system identification: learning from the theory of risk. *IFAC-PapersOnLine*, 48(28):1053–1058.

Vandenberghe, L. (2016). 4. Subgradients. http://www.seas.ucla.edu/~vandenbe/236C/lectures/subgradients.pdf. EE236C (Spring 2016). Accessed: 2018-03-14.