OPERATIONS RESEARCH

New Paradigms and Emerging Applications

Editors

Vilda Purutçuoğlu, Gerhard-Wilhelm Weber, and Hajar Farnoudkia

CRC Press
Taylor & Francis Group

A SCIENCE PUBLISHERS BOOK

Operations Research

New Paradigms and Emerging Applications

Editors

Vilda Purutçuoğlu
Professor, Department of Statistics
Middle East Technical University, Ankara, Turkey

Gerhard-Wilhelm Weber
Chairman, Marketing and Economical Engineering
Poznan University of Technology, Poznan, Poland

Hajar Farnoudkia
Department of Business Administration
Başkent University, Ankara, Turkey

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A SCIENCE PUBLISHERS BOOK

First edition published 2023
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

Library of Congress Cataloging-in-Publication Data (applied for)

ISBN: 978-1-032-34926-8 (hbk)
ISBN: 978-1-032-34929-9 (pbk)
ISBN: 978-1-003-32450-8 (ebk)

DOI: 10.1201/9781003324508

Typeset in Times New Roman
by Radiant Productions

Foreword

This book aims to provide Operation Research (OR) applications under twelve chapters organized in the form of four parts according to the main topics considered in these chapters. These four parts are optimization; data mining and clustering; business, science and finance; and medical application. Hereby, the continuous numbering of chapters is applied in those listed parts.

Accordingly, **Part 1** is allocated to Operation Research in optimization, given in three chapters.

Chapter 1 proposes a parameter estimation way for partially nonlinear problems which are semi-parametric regression models and the extension of the partially linear model that has gained importance in the statistical literature. These models are employed when the non-parametric regression model does not perform well. In this regard, firstly, nonlinear least square estimation is established based on the Taylor expansion of the nonlinear function. Then, a kernel-based bridge problem is employed to estimate the non-parametric component of the model. In the end, the optimization method is applied to choose the best estimation.

Chapter 2 provides a glimpse on the contributions and challenges towards more environmentally-friendly road traffic, by reviewing academic studies on how Operation Research has been used in controlling the complex transportation network since connected and analyzing traffic-related impacts, especially regarding environmental and air quality, of automated vehicles are not fully deployed yet on the roads. Operation Research in this matter is used to plan for future challenges and major impacts can be expected, as well.

Chapter 3 of this part represents some case studies for selecting suppliers and portfolio investment schemes by addressing the application of the multi-criteria method on discrete variables that are very important in complex decision-making problems.

Part II is devoted to new applications of Operations Research to Data Mining and Clustering and consists of three chapters. *Chapter 4* in this part is based on a survey about the dimension reduction methods such as clustering and principal component analysis (PCA). The main aim of this chapter is to investigate a semi-definite programming model that provides an effective solution to problems related to both PCA and clustering methods.

Chapter 5 of this part discusses some practical techniques for different types of clustering by using the formulation of the problems as an optimization method. Here, the clustering problems can be represented in terms of a real function of several real variables, and a set of arguments that give an optimal clustering.

Chapter 6, which is the last chapter of data mining and clustering portion, is dedicated to the application of these methods to meteorological data by presenting a review of the analysis. Then, as the illustration of the proposal approaches, the meteorological data are chosen. In the analyses, the data preparation and pre-processing are also explained in detail besides the clustering and the modeling in terms of the findings.

Part III of this book is based on operation research in business, science, and Finance in three distinct chapters. *Chapter 7* is about the foundations of market-making via stochastic optimal control. Market-making is a type of high-frequency trading that implies the quantitative trading of a short portfolio holding period. This chapter covers several results obtained by traditional techniques of stochastic optimal control related to market-making which is a significant component of financial research. *Chapter 8* is about the decision aid to drive the network with a better management of the system while there could be a conflict in used criteria such as the minimization of the cost and the maximization of the security simultaneously. Furthermore, there can be other objectives that may appear unexpectedly. So, this chapter is dedicated to explain why the formulation of the multi-objective network flow problem is a necessity and how it can be done.

Chapter 9 of the third part of this book is about operation research application in decision-making in finance focusing on the behavior of financial problems which are based on the investors' behavior introduced on sentiment. It also compares the forecasting performances of sentiments index by using different mathematical models.

Part IV represents operation research in medical application in three chapters. *Chapter 10* starts with a study covering necessary information about an algorithm and a stability approach for the acute inflammatory response dynamic model. Generally, the filamentary response wipes out the pathogens from the body and repairs the healthy case. Recently, mathematical models are being used to provide essential insights into the dynamics of the inflammatory response. On the other side, nonlinear dynamics have gained high importance in many areas that can describe the complicated conceptions within details. This study provides a numerical approximation to the complicated systems via nonlinear differential equations.

Chapter 11 titled Bayesian inference for the undirected network tries to estimate the conditional dependence between genes by using a Monte Carlo algorithm in case the number of parameters exceeds the number of observations. Here, the parameter in the precision matrix does not have a fixed dimension in each iteration. To overcome the problem, a reversible jump Markov chain Monte Carlo method, which is one of the OR algorithms optimizing both the fitted model to the data and the estimated parameters, is proposed. Furthermore, its alternatives such as Gibbs sampling, and Carlin Chibs methods, which are other advanced inference methods based on optimization of some score functions, are introduced. Moreover, as one of the powerful tools for modeling the biological network, the copula is introduced with its special type named the vine copula which tries to simplify the multivariate complicated model into the bivariate model. In the light of the vine copula, each undirected connection can be represented independently with the best copula family.

Finally, *Chapter 12*, the last chapter of the book, presents the comparison and transfer of the EMG data from two stations. In the compression of these signal data, two optimization techniques, namely, dynamic cosine transformation and principal component analysis, are implemented. Then, their performances are compared via distinct accuracy measures once their compressed values are classified by different clustering approaches that are well-known in the field of operation research.

As a result, we consider that this book can present the variety of applications of operation research techniques in different problems. We hope that the references used in each part can be also useful for the new researchers while deeply learning the theoretical aspects of the selected methods. Hence, we hope that the book can open new avenues for novel researches.

Preface

This original book on Operational Research: New Paradigms and Emerging Applications surveys and details newest technology in analytics and intelligence computing, artificial intelligence (AI) and operational research (OR) which are reducing the dimensions of data coverage as well as variables worldwide. This compendium discusses code of intelligent optimization which can be applied in various branches of optimization, data mining and clustering, economic/finance and medical applications. Involving modern and emerging techniques of OR and AI, and applying them together with strong and evolutionary algorithms to real-life problems for strategical, but also daily applications, this compendium elaborates all areas of OR results, methods and applications. By the rich diversity of this handbook, the state-of-the-art developments in quickly advancing key technologies are covered. In this way, with our reference work we hope to be useful for students and emerging scientists of engineering and science, management and economics, social science and the art, for researchers and scholars who are employed in OR supported industries, for decision-makers and designers of tomorrow's World.

We editors hope that the chosen subjects and picked areas reflect a core sample of international OR research facing emerging, challenging, complex and even long-enduring problems of our environment and their field in economics and finance, natural sciences and engineering, healthcare and medicine, industry and city planning, through the results and tools of OR-Analytics. We are very grateful to the publishing house of CRC Press for the honor of accommodating this front running project in intelligence and science, operations and implementation. We convey particular thanks to the directors, editors and managers of CRC Press as well as to its editorial management and team, for their steady interest, care and encouragement, recommendations, support and guide in every respect. We thank our respected authors by their diligent work and readiness to share their newest findings, insights and results with over international community. Now we hope that the research of our authors collected and edited by us will be an inspiration for cooperation and joint implementation, improvement and friendship at the global stage and premium level.

Acknowledgement

I would like to thank deeply to Prof. Dr. Gerhard-Wilhelm Weber for his great support throughout my academic career, and Dr. Hajar Farnoudkia for her hardworking and willingness in this book. Finally, my utmost sincerest thanks go to my parents, my mother Yurdagül Purutçuoğlu, my father Ahmet Purutçuoğlu, my sister Prof. Dr. Eda Purutçuoğlu and my sweat daughter Vera to their big love and support.

—Vilda Purutçuoğlu

I would like to cordially thank Prof. Dr. Vilda Purutçuoğlu for her vision, care and inspiration, and Dr. Hajar Farnoudkia for her dedicated hard-work. All of these made this book become so beautiful and such a success.

—Gerhard-Wilhelm Weber

I want to thank my PhD advisor, Prof. Dr. Vilda Purutçuoğlu for her passion and knowledge who taught me how to study, search and write. I owe a big part of my success to her. Also I think here is a good opportunity to thank Assoc. Prof. Dr. Babek Erdebill who encouraged and helped me to be one of the academic world members.

—Hajar Farnoudkia

All editors also appreciate the great contribution of all authors as they accepted to share their knowledge and researches with them. Finally, they like to convey their sincerest thanks to Directors, Editors, Managers of CRC Press and Vijay Primlani, the acquisitions editor of CRC Press, and his team for their supports and guide in every respect.

Contents

Part III: Operation Research in Business Science and Finance

Part IV: Operation Research in Medical Application

Contributors

Akan, Erhan
Turkish Aerospace Industries
Ankara, Turkey

Aktaş Dinçer, Hayriye
Fatih Sultan Mehmet Vakıf University
İstanbul, Turkey

Arı, Fikret
Ankara University
Ankara, Turkey

Bandeira, Jorge
University of Aveiro
Aveiro, Portugal

Batmaz, İnci
Middle East Technical University
Ankara, Turkey

Belkacem, Salima Nait
University of M'hamed Bougara
Boumerdès, Algeria

Can Erkuş, Ekin
Middle East Technical University
Ankara, Turkey

Farnoudkia, Hajar
Başkent University
Ankara, Turkey

Farzin Asanjan, Mahdieh
Middle East Technical University
Ankara, Turkey

Freitas, Adelaide
University of Aveiro
Aveiro, Portugal

Gökçay, Didem
21yy Ltd.
Ankara, Turkey

Gökçe, Aytül
Ordu University,
Ordu, Turkey

Gomez-Rueda, Mario Sergio
Universidad Pontificia Bolivariana
Medellín, Colombia

Gürbüz, Burcu
Johannes Gutenberg-University
Mainz, Mainz, Germany

İleri, Fatih
Turkish Aerospace Industries
Ankara, Turkey

Kalaycı, Betül
Middle East Technical University
Ankara, Turkey

Lotero, Laura
Universidad Pontificia Bolivariana
Medellín, Colombia

Macedo, Eloisa
University of Aveiro
Aveiro, Portugal

Purutçuoğlu, Vilda
Middle East Technical University
Ankara, Turkey

Savku, Emel
University of Oslo
Oslo, Norway

Somuncuoğlu, Abdullah Nuri
Middle East Technical University
Ankara, Turkey

Taylan, Pakize
Dicle University
Diyarbakır, Turkey

Tchemisova, Tatiana
University of Aveiro
Aveiro, Portugal

Wilhelm Weber, Gerhard
Poznan University of Technology
Poznan, Poland

Yerlikaya Özkurt, Fatma
Atılım University
Ankara, Turkey

I

OPERATION RESEARCH IN OPTIMIZATION

Chapter 1

Kernel Based C-Bridge Estimator for Partially Nonlinear Model

Pakize Taylan

1.1 Introduction

Semi-parametric regression models (SPRMs) [28] deal with regression models which consider the effects of both the parametric and nonparametric regression models simultaneously. They are very helpful for data analysis since they keep the flexibility of nonparametric models and the properties easy interpretation of parametric models to comment baseline function f. Therefore, SPRMs have attracted considerable attention in recent years and have been studied by many researchers interested in data analysis. For this chapter, we consider partially nonlinear models (PNLMs) that are SPRMs and extensions of partially linear models (PLMs) [15] which have been popular in the statistical literature. PNLMs are employed when nonparameric regression does not perform well.

A standard form of partially nonlinear model is defined as

$$Y_i = f(X_i, \delta) + h(U_i) + \varepsilon_i, i = 1, 2, ..., n, \tag{1.1}$$

where $X_i = (x_{i1}, x_{i2}, ..., x_{ip})(p \geq 1)$ and $U_i = (u_{i1}, u_{i2}, ..., u_{iq})(q \geq 1)$ are considered as the vectors of independent and identically distributed explanatory variables, respectively, Y_i is the response variable for the ith case, $\delta =$

Dicle University.

$(\beta_1, \beta_2, ..., \beta_p)^T$ is unknown parameters' vector, $f(.,.)$ is a pre-assigned function, h is an unknown smooth function from o^q to o^l, and $\varepsilon_i (i = 1, 2, ..., n)$ are independent random errors with $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$, respectively.

PNLMs are widely used in the literature due to their usefulness mentioned above. Li and Mei in 2013 [21] developed new estimation procedures for parameters in the parametric component and they formed consistency and asymptotic normality of the estimator which they achieved. Also, they proposed estimation procedures that consider a generalized F test for the nonparametric component in the PNLMs. Severini and Wong in 1992 [30] introduced a geometric framework that contains the concept of the least favorable curve for PNLMs [30]. Zhong et al. in 2000 [37] established three types of developed approximate confidence regions for the parameter in terms of curvatures for PNLMs, considering Severin's geometric framework. Application of the finite series approximation method to a partially nonlinear model and its some new results were handled by Xie et al. in 1997 [13]. Hung and Chen in 2008 [18] studied the parameter estimation problem for the nonlinear partial spline model, $f(X, \delta) = \delta^T X$ that is a special form of PNLMs when a nonparametric component is approximated by some graduating function. Wang and Ke in 2009 [34] developed an estimation problem for smoothing spline semi-parametric nonlinear regression models by considering the penalized likelihood and they solved it by Gauss-Newton and back-fitting algorithms. The parameters in the problem were estimated by employing generalized cross-validation(GCV) [8] and generalized maximum likelihood methods (GML) [36].

This chapter proposes an estimation procedure for PNLMs where its nonparametric component $h(U_i)$ is considered as an additive nonparametric component [6, 32]. To achieve an estimation of parameter for both nonparametric and parametric parts, firstly, we establish a nonlinear least square estimation problem based on the Taylor expansion of nonlinear function $f(.)$ at initial value $\hat{\delta}_c$ where $\hat{\delta}_c$ is a consistent estimate of δ. Secondly, we establish a kernel based bridge problem to estimate the nonparametric component of PNLMs, say $h(.)$. Then, we solve the problem that we established with the famous method of convex optimization called conic quadratic programming (CQP) [5]. Finally, for each δ we evaluate $\hat{\theta}$ and hence, h in kernel based bridge problem that gives a new curve over h. Then, we produce a solution profile for h and choose the $\hat{\delta}$ which is the minimum over all these curves as an estimation of δ.

The remainder of this chapter is formed as follows. The second Section includes an introduction of the bridge estimator for the regression model. Additive approximation and kernel based bridge estimation problem for PNLMs are handled in the third section. The fourth section presents the construction of the estimation problem established in the third section as a CQP problem in order to use the advantages of CQP. Finally, the fifth section gives a short conclusion.

1.2 Bridge Estimators

Penalized estimation methods such as penalized linear least squares and penalized likelihood, have drawn much attention in recent years, and it has been used quite a lot by many researchers because they present a method for selecting of variables and estimating of parameters simultaneously in linear regression given as

$$y_i = x_i^T \delta + \varepsilon_i, i = 1, 2, ..., n. \tag{1.2}$$

Here, $y_i \in o$ is a i-th response variable, $x_i (i = 1, 2, ..., n)$ is a p-vector of covariates, δ is a p-vector of unknown parameters, and ε_i is an n-vector of identically distributed, independent random errors. The bridge estimation method suggested by Frank and Friedman in 1993 [11] consists of a large class of the penalty methods considering penalty function $\sum |\delta_j|^\alpha$ with $\alpha > 0$. Bridge estimator, $\hat{\delta}_B$, can be determined by solving optimization problem given as

$$\text{minimize}_\delta \sum_{i=1}^{n} (y_i - x_i^T \delta)^2 + \varphi \sum_{j=1}^{p} |\delta_j|^\alpha, \tag{1.3}$$

where $|.|$ is the L_2-norm of the vector, φ is a penalty parameter that provides a trade off between the first and the second term. As seen in Equation 1.3, the objective function is penalized by the L_α-norm to obtain bridge estimator $\hat{\delta}_B$ and it shrinks the estimates of the parameters in Equation 1.2 towards 0. Liu et al. in 2007 [22] discussed the effect of the L_α penalty with different cases of α. If $\alpha = 1$, bridge estimation produces Lasso (Least Absolute Shrinkage Operator) [32]; if $\alpha = 2$, it produces Ridge or Tikhonov regularization [17] estimation. For $\alpha \leq 1$, the bridge estimator manages to select significant variables for the regression model by shrinking small $|\delta_j|$s to exact zeros. However, Knight and Fu in 2000 [20] handled the asymptotic distributions of bridge estimators when the number of covariates is fixed and they noted that the amount of shrinkage towards zero increases with the magnitude of the regression coefficients being estimated in case of $\alpha > 1$.

Also, Liu et al. in 2007 [22] pointed out that in penalized estimation problem, to obtain acceptable bias for large parameters, the value of α is not chosen too high than necessary.

They give the following example for $\alpha \in [0, 2]$ to explain this situation.

Liu et al. in 2007 [22] considered a simple linear regression model with one parameter δ and one observation $y = \delta + \varepsilon$, where ε is a random error with mean zero and variance σ^2 for illustrating the effect of L_α penalties with respect to different α. As a result of this example, they made the following inferences:

i) Bridge solution is ridge $\hat{\delta} = y/(\varphi + 1)$ in case $\alpha = 2$ and it is biased with $Var(\hat{\delta}) = Var(y/(\varphi + 1)^2)$. Hence, $\hat{\delta}$ is better than y when the bias is smaller compared to variance deduction.

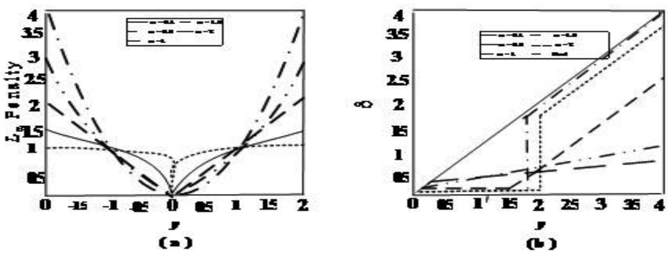

Figure 1.1: **a**: Plots of L_α penalties for different α, **b**: The solutions $\hat{\delta} = \text{argmin} F_\alpha(\delta)$ with respect to the L_α penalties in **(a)** with $\varphi = 3$ where $F_\alpha(\delta) = (\delta - y)^2 + \varphi|\delta|^\alpha$ [22].

ii) Bridge solution is a Lasso solution $\hat{\delta} = sgn(y)[|y| - \varphi/2]$ that gives a thresholding rule, since small $|y|$ leads to a zero solution.

iii) They conclude that $\hat{\delta} = 0 \iff \varphi > |y|^{2-\alpha}\left(\frac{2}{2-\alpha}\right)\left[\frac{2(1-\alpha)}{2-\alpha}\right]^{1-\alpha}$ in case $\alpha \in (0,1)$, that is, $|y| < \left[\varphi\left(\frac{2-\alpha}{2}\right)\left(\frac{2-\alpha}{2(1-\alpha)}\right)^{1-\alpha}\right]^{1/(2-\alpha)}$ [20].

As pointed out in [24], L_α penalty in the problem Equation 1.3 is strictly convex when $\alpha > 1$ and strictly non-convex when $\alpha < 1$. When $\alpha = 1$, it is still convex, but, not differentiable at the origin. It is clearly shown that the elasticnet penalty is between $\alpha = 1$ (Lasso) and $\alpha = 2$ (ridge), and it is strictly convex. As a result, when $\alpha \geq 1$, the problem Equation 1.3 is convex and solvable without using the approximation. Therefore, we handle bridge estimators for partially nonlinear model by using advantages of convex optimization, especially, conic quadratic programming [5].

1.3 PNLMs with Additive Approximation and Bridge Estimation

1.3.1 Construction of additive nonparametric component

In this section, we present the form of the bridge penalty for PNLM. Let us consider $\{(y_i, x_i, u_i), i = 1, 2, \ldots, n\}$ a random sample from model expressed by Equation 1.4. The nonlinear least squares objective function for Equation 1.4 is written as

$$Q(h, \delta) = \sum_{i=1}^{n} \left(y_i - f(x_i, \delta - h(u_i)\right)^2. \tag{1.4}$$

Estimation procedure for parameters in the PNLM consists of two-step through Equation 1.4. In the first step, a linear approximation of $f(., \delta)$ is taken into consideration through Taylor expansion since $f(., \delta)$ is nonlinear with respect to δ. Then, we try to find the minimizer of Equation 1.4 by solving normal equations obtained from the derivative of $Q(h,\delta)$ concerning δ. However, normal equations cannot be solved analytically due to their non-linearity; therefore, iterative techniques such as the Newton-Raphson algorithm should be used [3]. By applying Taylor expansion to $f(.,\delta)$, at $\hat{\delta}_c$ where $\hat{\delta}_c$ is a consistent estimate of δ as an initial point, thus, we get

$$f(x,\delta) = f(x,\hat{\delta}_c) + f'(x,\hat{\delta}_c)^T \left(\delta - \hat{\delta}_c\right) + o_p\left(\| \delta - \hat{\delta}_c \|\right). \tag{1.5}$$

The initial point $\hat{\delta}_c$ in Equation 1.5 is obtained from the solution of the following nonlinear least squares optimization problem:

$$\hat{\delta}_c = \text{argmin}_\delta \sum_{i=1}^{n} \left(y_{(i+1)} - y_{(i)} - f(x_{(i+1)},\delta) + f(x_{(i)},\delta)\right)^2, \tag{1.6}$$

where $\left(x_{(i)}, t_{(i)}, y_{(i)}\right)$ $(i = 1,2,...,n)$ is an ordered sample from the smallest to the largest according to the value of the variable u_i [35]. Li and Mei in 2013 [21] have shown that under some conditions, $\hat{\delta}_c$ is root n consistent. Thus, the ith sample for response variable Y_i, y_i can be written as

$$y_i = f(x_i,\hat{\delta}_c) + f'(x_i,\hat{\delta}_c)^T \left(\delta - \hat{\delta}_c\right) + h(u_i) + \varepsilon_i. \tag{1.7}$$

Let $z_i = y_i - f(x_i,\hat{\delta}_c) + f'(x_i,\hat{\delta}_c)^T f(x_i,\hat{\delta}_c)$. Then, we get the following linear approximation model,

$$z_i = f'(x_i,\hat{\delta}_c)^T \delta + h(u_i) + \varepsilon_i \tag{1.8}$$

or in matrix

$$z = F\delta + h(u) + \varepsilon, \tag{1.9}$$

where $z = (z_1, z_2, ..., z_n)^T$ is n-vector of adjusted response variables, F is $(n \times p)$-dimensional derivative matrix whose (i, j) element is $\partial f(x_i,\delta)/\partial \delta_j|_{(\delta=\hat{\delta}_c)}$ and $h(u)$ is n-vector of regression function $h(u_i)$. In the second step, available estimation methods can be directly used to estimate δ by considering the partial linear model given as Equation 1.9.

Here, we prefer the profile least squares technique which is also a two-step process. In the first step, for a given δ let $y_i = z_i - f'(x_i,\hat{\delta}_c)^T \delta$. Then, we rewrite model Equation 1.4 as

$$\nu = h(u) + \varepsilon, \tag{1.10}$$

where $\nu = (\nu_1, \nu_2, ..., \nu_n)^T$ can be considered sample from model Equation 1.10 and $h(u) = (h(u_1), h(u_2), ..., h(u_n))^T$ is considered n-vector for unknown regression function. The model Equation 1.10 is a q-dimensional nonparametric regression model. Hence, $h(.)$ may be estimated by one of the nonparametric estimation methods such as smoothing spline [10], k-nearest-neighbors [16], kernel estimation, [29] and local least squares estimation [23] for characterizing nonlinear trend in the model.

The large number of explanatory variables $U_j(j = 1,2,...,q)$ in the model Equation 1.10 causes an increase in the variance of nonparametric estimators and therefore, an increase in test error. This situation, called the curse of dimensionality [4], makes the interpretations of the model very difficult and prevents obtaining reliable results. These challenges of nonparametric regression were overcome with *Additive Models* recommended by Stone in 1985 [32]. Additive models [32] offer estimates that have lower variance than nonparametric models, and can have a lower bias than parametric ones. In additive approximation, the change in the response variable corresponding to each explanatory variable is explained by the estimation of individual terms and it employs univariate smoothing. Therefore, the additive regression model will be considered for estimation of smooth function $h(.)$ to obtain a profile nonlinear least square [21] estimate of δ. Given n realizations for pairs $(u_i, v_i)(i = 1,2,...,n)$ with each $u_i = (u_{i1},...,u_{iq})$, the additive model for $h(u_i)$ is written as

$$E(V_i|u_{i1},...,u_{iq}) = h(u_i) = \delta_0 + \sum_{j=1}^{q} h_j(u_{ij}), i = 1,...,n, \tag{1.11}$$

under the assumption $E(h_j(u_{ij})) = 0$ in order to avoid a different intercept in each h_j function [6, 32]. The additive estimate of each function, $h(u_i)$ by considering Equation 1.11 is obtained by iteration scheme, called back-fitting algorithm, which is proposed by Friedman and Stuetzle in 1981 [14].

In this study for additive approximation of function $h(u)$, the functions h_j will be considered as spline functions, that is, linear combination of the parametrical form

$$h_j(u) = \sum_{l=1}^{d_j} \theta_l^j g_l^j(u), \tag{1.12}$$

where $g_l^j : R \longrightarrow R$ is the l-th transformation (base spline) of $u, (l = 1,2,...,d_j)$, θ_l^j is the (l, j)-th entry of the family

$$\theta = \left(\theta_l^j\right)$$

in which $l = 1,...,d_j$ and $j = 1,...,q$ as well as for the sake of simplicity, by introducing additional terms with coefficients 0, we may assume that $g_l^j \equiv g_l$, $d_j \equiv d, (j = 1,2,...,q)$ such that the family becomes a matrix.

Natural spline [26], B-spline [9] and multivariate adaptive regression spline [12] are examples of spline functions commonly used in data analysis.

Let us now explicitly insert the parametrical form Equation 1.12 of the functions h_j into Equation 1.11. Then, Equation 1.11 looks as follows:

$$E(V_i|u_{i1},...,u_{iq}) = h(u_i) = \delta_0 + \sum_{j=1}^{q}\sum_{l=1}^{d}\theta_l^j g_l(u_{ij}), i = 1,...,n. \tag{1.13}$$

For all $i = 1,2,...,n$ we can write

$$\begin{aligned}&\textstyle\sum_{j=1}^{q}\sum_{l=1}^{d}\theta_l^j g_l(u_{ij}) = \theta_1^1 u_1(u_{ij}) + ... + \theta_d^1 u_d(u_{ij}) + ... + \theta_1^q u_1(u_{iq}) + ... + \theta_d^q u_d(u_{iq})\\ &= (g_1(u_{i1},...,g_d(u_{i1}))\left(\theta_1^1,...,\theta_d^1\right)^T + ... + (g_1(u_{im},...,g_d(u_{im}))\left(\theta_1^m,...,\theta_d^m\right)^T\end{aligned} \tag{1.14}$$

or

$$\sum_{j=1}^{q}\sum_{l=1}^{d}\theta_l^j g_l(u_{ij}) = G_i^1\theta_i^1 + ... + G_i^q\theta_i^q + \left(G_i^1,...,G_i^q\right)\left(\theta_i^{1T},...,\theta_i^{qT}\right)^T = G_i\theta \tag{1.15}$$

where $\theta^j := \left(\theta_1^j,...,\theta_d^j\right)^T$, $\theta = \left(\theta^{1T},...,\theta^{qT}\right)$, $G_i^j := (g_1(u_{ij}),...,g_d(u_{ij}))$ and $G_i := \left(G_i^1,...,G_i^m\right)$ $(i = 1,2,...,n)$. If Equation 1.15 is used in Equation 1.11 and assuming that δ_0 is fixed via the estimation $\hat{\delta}_{a0} := ave(i|i = 1,2,...,n)$ by the arithmetic mean of the values i, then, Equation 1.11 is obtained $\hat{\delta}_0$ as

$$h(u_i) = \hat{\delta}_0 + G_i\theta, i = 1,...,n. \tag{1.16}$$

Hence, as a result of this last equation, Equation 1.10 turns into the following form:

$$v = \hat{\delta}_0 1 + G\theta + \varepsilon \tag{1.17}$$

where 1 is a n-vector of ones, G is an $(n \times qd)$-dimensional matrix with ith row $G_i := \left(G_i^1,...,G_i^m\right)$.

1.3.2 Kernel based bridge estimation for PNLMs

To obtain bridge estimation of the parameters θ in Equation 1.17, the objective function is defined as

$$L_B(\hat{\delta}_0, \theta_l^1,...,\theta_d^q) := \sum_{i=1}^{n}\{v_i - \hat{\delta}_0 - G_i\theta\}^T + \varphi\sum_{j=1}^{q}\sum_{l=1}^{d}|\theta_l^j|^\alpha, \tag{1.18}$$

which is just a penalized least square objective function penalized by L_α-norm. This penalization provides shrinking of the estimates of the parameters (θ_l^j) in Equation 1.17 towards 0. Here, $\varphi > 0$ is a penalty or smoothing parameter that provides a trade-off between the goodness of data fitting expressed by the first sum and the penalty function expressed with the second sum. As can be seen, the smoothing parameter φ influences the smoothness of a fitted curve. Therefore, it should be estimated by one of the well known methods such as generalized cross validation (GCV) [8], Akaike information criteria (AIC) [7] and minimization of an unbiased risk estimator (UBRE) [8]. The goal of smoothness is sometimes also called stability, robustness, or regularity. In fact, in the theory of inverse problems one wants to guarantee that the estimation is sufficiently stable with respect to noise and other forms of perturbation.

The bridge estimation of θ, $\hat{\theta}_B$ is obtained by solution of the optimization problem

$$\text{minimize}_{\theta} L\left(\hat{\delta}_0, \theta_l^1, ..., \theta_d^q\right). \tag{1.19}$$

To solve problem Equation 1.18, we consider the kernel estimation method [29] developed for modelling strong non-linearity between independent and dependent variables. This method provides estimates of the regression function by stating the nature of the local neighborhood expressed by a *kernel function* $K_\lambda(x_0, x)$, and the nature of the class of regular functions fitted locally. In this sense, a transformation of the original data is used through kernel functions which are considered as weights, and they form a kernel matrix to produce weighted average estimators. Thus, by using this method, the complexity of the calculations is considerably reduced since the model is performed by considering the kernel matrix that summarizes the similarity in the observation values instead of the original data.

The simplest form of kernel estimate is the Nadaraya-Watson weighted average [29].

$$\hat{h}(u_0) = \frac{\sum_{i=1}^{n} K_\lambda(u_0, u_i)\nu_i}{\sum_{i=1}^{n} K_\lambda(u_0, u_i)}, \tag{1.20}$$

where $K_\lambda(u_0, u_i) := K(\| u_i - u_0 \|_2 / \lambda$ a kernel function defined as $K : o \longrightarrow o$, providing

$$\int K(u)du = 1, K(-x_0) = K(x_0). \tag{1.21}$$

The most typical kernels functions for $u_i \in o^q$ are (i) Uniform, (ii) Epanechnikov, (iii) Gaussian, Quartic (biweight), and (iv) Tricube (triweight) given as follows:

i $K(u) = \frac{1}{2} 1_{(\|u\| \leq 1)0}$

ii $K(u) = \frac{3}{4}(1 - \| u \|^2) 1_{(1-\|u\| \leq 1)0}$

iii $K(u) = \frac{15}{16}(1 - \| u \|^2) 1_{(1-\|u\| \leq 1)0}$

iv $K(u) = \frac{35}{32}(1 - \| u \|^2)^3 1_{(1-\|u\| \leq 1)0}$

Hence, the bridge estimate of θ in Equation 1.17 based on a kernel function can be obtained by minimizing the penalized residual sum of squares given

$$L_{KB}(\hat{\delta}_0, \theta_l^1, ..., \theta_d^q) := \sum_{i=1}^{n} K_\lambda(u_0, u_i)\{\nu_i - \hat{\delta}_0 - G_i\theta\}^2 + \varphi \sum_{i=1}^{q} \sum_{l=1}^{d} |\theta_l^j|^\alpha, \tag{1.22}$$

where dimensions of all matrices and all vectors are the same as the corresponding vectors and matrices in Equation 1.22. If we take $b_i = v_i - \hat{\delta}_0 (i = 1, 2, ..., n)$, $b = (b_1, ..., b_n)^T, A = K^{1/2}G, a = K^{1/2}b$ where $K^{1/2}$ is $(n \times n)$-dimensional diagonal matrix with ith diagonal element $[L_\lambda(u_0, u_i)]^{1/2}$, then, kernel based penalized residual sum of squares looks as

$$\begin{aligned} L_{KB}(\hat{\delta}_0, \theta_l^1, ..., \theta_d^q): &= \textstyle\sum_{i=1}^n K_\lambda(u_0, u_i)\{v_i - \hat{\delta}_0 - G_i\theta\}^2 + \varphi \sum_{i=1}^q \sum_{l=1}^d, |\theta_l^j|^\alpha. \\ &= \textstyle\sum_{i=1}^n K_\lambda(u_0, u_i)\{b_i - G_i\theta\}^2 + \varphi \sum_{i=1}^q \sum_{l=1}^d, |\theta_l^j|^\alpha. \\ &= (G\theta - b)^T K_\lambda (G\theta - b) + \varphi \textstyle\sum_{i=1}^q \sum_{l=1}^d, |\theta_l^j|^\alpha. \\ &= \| A\theta - a \|^2 + \varphi \textstyle\sum_{i=1}^q \sum_{l=1}^d, |\theta_l^j|^\alpha. \end{aligned} \tag{1.23}$$

The kernel based bridge estimation of θ, $\hat{\theta}_{KB}$ is obtained by solution of the optimization problem

$$\text{minimize}_\theta L(\hat{\delta}_0, \theta_l^1, ..., \theta_d^q). \tag{1.24}$$

We handle two well known special cases of kernel based bridge estimator. When $\alpha = 2$, the kernel based bridge estimator,$\hat{\theta}_{KB}$ will be equivalent to the kernel based ridge estimator $\hat{\theta}_{KB}^R$ [17] and it is obtained from the solution of the following problem:

$$\text{minimize}_\theta \| A\theta - a \|_2^2 + \varphi \| \theta \|_2^2, \tag{1.25}$$

where $\| 0 \|_2$ stands for Euclidean norm.

The optimization problem given in Equation 1.24 can be solved by using SVD of coefficient matrix A [2] and its solution is

$$\hat{\theta}_{KB}^R = \left(A^T A + \varphi\right)^{-1} A^T a. \tag{1.26}$$

For a suitable value of φ, the ridge estimator has a smaller mean squared error than that of the least squares estimator. When $\alpha = 1$, the kernel based bridge estimator, $\hat{\theta}_{KB}$ will be equivalent to the kernel based Lasso estimator, $\hat{\theta}_{KB}^{Lasso}$ that is the solution of the following problem:

$$\text{minimize}_\theta \| A\theta - a \|_2^2 + \varphi \| \theta \|_1 \tag{1.27}$$

where $\| \theta \|_1 := \sum_{j=1}^q \sum_{l=1}^d |\theta_l^j|^\alpha$. If some components of T are 0, the objective function in problem Equation 1.27 will be non-differentiable, so the problem cannot be solved by standard unconstrained optimization methods.

In this section, we consider a convex optimization method called conic quadratic programming (CQP) to solve problem Equation 1.22, as they provide computationally easy study and theoretically efficient solutions, and this solution will be named C_K-bridge estimation.

1.4 On Conic Optimization and Its Application to Kernel Based Bridge Problem

1.4.1 Convex and conic optimization

Convex optimization [5] is a special class of mathematical optimization problems such as least-squares and linear programming, and it handles problems aiming to minimize a convex function based on a convex set. The great advantage of expressing a problem as a convex optimization problem is to obtain a reliable and effective solution employing interior-point methods. Convex optimization programs have been employed for many years in scientific research including both theory and practice, as they present a strong theory of duality that has a very interesting comment in terms of the original problem. It has also found wide application in combinatorial optimization and global optimization, as it is capable of finding optimal value bounds as well as its approximate solution for optimization problems. Convex optimization contains different important classes of optimization problems such as CQP which is considered for our problem, semidefinite programming, and geometric programming. Brief information related with CQP will be given in the following by benefiting from [5].

A CQP is a conic problem

$$\text{minimize}_{\phi} c^T \phi, \quad \text{where} \quad S\phi - s \in C. \tag{1.28}$$

Here, the cone C consists of direct product of *"Lorentz cones"* defined as

$$L^{n_i+1} = \{\phi = (\phi_1, \phi_2, ..., \phi_{n+1})^T \in R^{n_i+1} | \phi_{n_i+1} \geq \sqrt{\phi_1^2 + \phi_2^2 + ... + \phi_{n_i}^2}\} (n_i \geq 1, n_i \in Y). \tag{1.29}$$

The geometric interpretation of a quadratic (or second-order) cone is shown in Figure 1.2 for a cone with n_i variables, and illustrates how the boundary of the cone resembles an ice-cream cone. The 1-dimensional quadratic cone simply states non-negativity $\phi_{n_i+1} \geq 0$. More generally, partitioning the data matrix $[S_i; s_i]$ by

$$[S_i; s_i] = \begin{bmatrix} D_i & d_i \\ p_i^T & q_i \end{bmatrix},$$

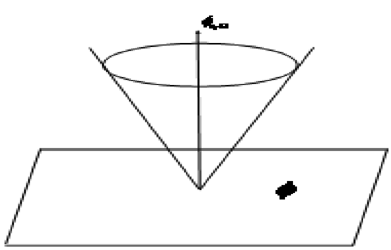

Figure 1.2: Boundary of quadratic cone $\phi_{n_i+1} \geq \sqrt{\phi_1^2 + \phi_2^2 + ... + \phi_{n_i}^2}$.