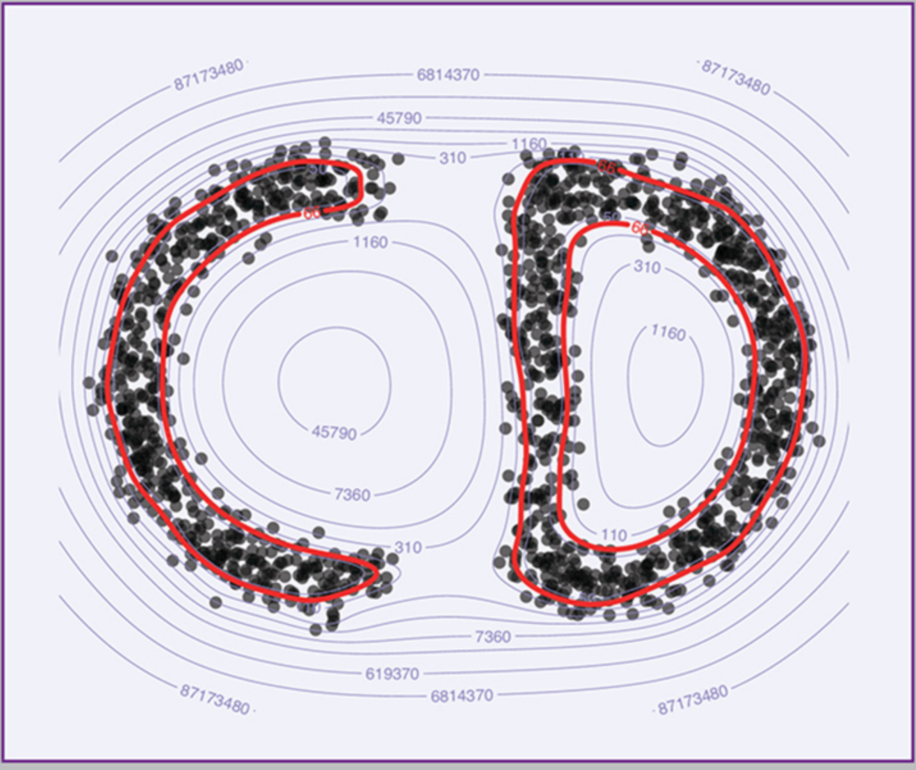
Cambridge Monographs on Applied and Computational Mathematics

The Christoffel–Darboux Kernel for Data Analysis

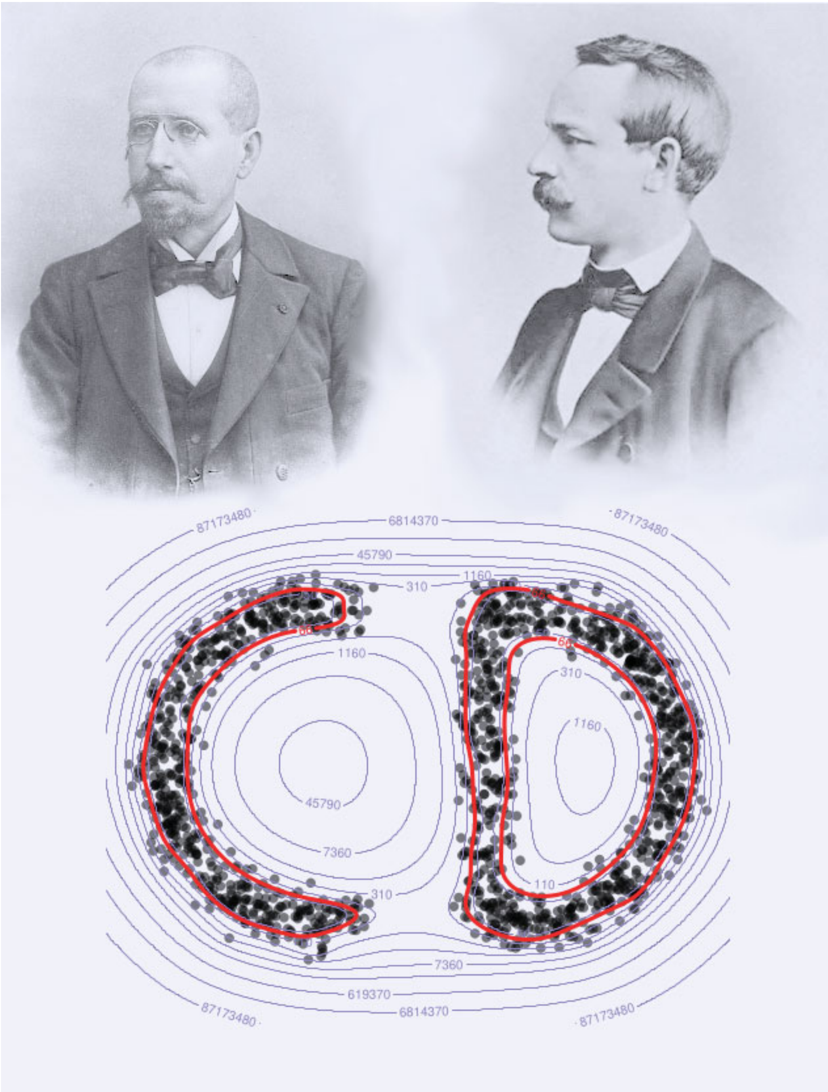
Jean Bernard Lasserre, Edouard Pauwels and Mihai Putinar



CAMBRIDGE MONOGRAPHS ON APPLIED AND COMPUTATIONAL MATHEMATICS

38

The Christoffel–Darboux Kernel for Data Analysis


87173480
6814370
87173480
45790
1160
310
60
1160
60
310
1160
45790
7360
310
110
7360
619370
87173480
6814370
87173480

The *Cambridge Monographs on Applied and Computational Mathematics* series reflects the crucial role of mathematical and computational techniques in contemporary science. The series publishes expositions on all aspects of applicable and numerical mathematics, with an emphasis on new developments in this fast-moving area of research.

State-of-the-art methods and algorithms as well as modern mathematical descriptions of physical and mechanical ideas are presented in a manner suited to graduate research students and professionals alike. Sound pedagogical presentation is a prerequisite. It is intended that books in the series will serve to inform a new generation of researchers.

A complete list of books in the series can be found at
www.cambridge.org/mathematics.

Recent titles include the following:

22. The mathematical foundations of mixing, *Rob Sturman, Julio M. Ottino & Stephen Wiggins*
23. Curve and surface reconstruction, *Tamal K. Dey*
24. Learning theory, *Felipe Cucker & Ding Xuan Zhou*
25. Algebraic geometry and statistical learning theory, *Sumio Watanabe*
26. A practical guide to the invariant calculus, *Elizabeth Louise Mansfield*
27. Difference equations by differential equation methods, *Peter E. Hydon*
28. Multiscale methods for Fredholm integral equations, *Zhongying Chen, Charles A. Micchelli & Yuesheng Xu*
29. Partial differential equation methods for image inpainting, *Carola-Bibiane Schönlieb*
30. Volterra integral equations, *Hermann Brunner*
31. Symmetry, phase modulation and nonlinear waves, *Thomas J. Bridges*
32. Multivariate approximation, *V. Temlyakov*
33. Mathematical modelling of the human cardiovascular system, *Alfio Quarteroni, Luca Dede' Andrea Manzoni & Christian Vergara*
34. Numerical bifurcation analysis of maps, *Yuri A. Kuznetsov & Hil G.E. Meijer*
35. Operator-adapted wavelets, fast solvers, and numerical homogenization, *Houman Owhadi & Clint Scovel*
36. Spaces of measures and their applications to structured population models, *Christian Düll, Piotr Gwiazda, Anna Marciniak-Czochra & Jakub Skrzeczkowski*
37. Quasi-interpolation, *Martin Buhmann & Janin Jäger*

The Christoffel–Darboux Kernel for Data Analysis

JEAN BERNARD LASSERRE
LAAS-CNRS, University of Toulouse, CNRS

EDOUARD PAUWELS
IRIT, University of Toulouse, CNRS

MIHAI PUTINAR
University of California, Santa Barbara

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108838061
DOI: 10.1017/9781108937078

First published 2022

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-83806-1 Hardback

Contents

Foreword

Francis Bach, INRIA

Characterizing and analyzing probability distributions through their moments has a long history in probability, statistics, optimization, signal processing, machine learning and all related fields. Christoffel–Darboux (CD) kernels are well-studied mathematical objects which were originally introduced for very different purposes. They turn out to benefit from interesting properties within a moment-based analysis context leading to interesting applications.

The theory of CD kernels and their relationship to moments is more than a century old, and is still an active field of mathematical research. The motivations for studying such objects arose from fundamental mathematics, with orthogonal polynomials and approximation theory, and have remained quite disconnected to applied disciplines centered on inference from data. Yet CD kernels turn out to have several appealing properties from an empirical inference perspective. They can be defined from moments, requiring only conceptually simple numerical operations. Furthermore, theory shows that many subtle properties of the underlying distribution can be obtained from the CD kernels of increasing orders, such as its support.

This book demonstrates the potential of CD kernels as an empirical inference tool in a data analysis context. It investigates the consequences of the favorable properties of CD kernels in a statistical context where one only has access to empirical measures and empirical moments. This original thematic positioning naturally leads to questions at the interface between applied statistical inference and the CD kernel literature. These include statistical connections between empirical Christoffel functions and its large-sample limit, quantitative estimates and bounds, and consequences for applications.

Interestingly, the Christoffel–Darboux kernel is a reproducing kernel for a space of polynomials, a notion that is now common in statistics and machine learning. The construction is, however, very different with a sample dependency of underlying scalar product and norm, which are adapted to the empirical measure. This contrasts with more usual machine learning applications where the scalar product is fixed and given, and provides an efficient basis for polynomial estimation with a natural interpretation for increasing degree orders.

Many aspects of the theory of Christoffel functions and the associated Christoffel–Darboux kernels are well established and have become classical in the polynomial approximation literature. This book provides a unified and clear exposition of the main tools and algorithms, with a strong focus on data analysis applications. It shows in particular how the new polynomial kernels can be efficiently used for many relevant tasks, such as support estimation, outlier detection or experimental design.

Symbols

$\mathbb{N}$	the set of natural numbers
$s(d)$	$\binom{n+d}{n}$
$\mathbb{Z}$	the set of integers
$\mathbb{Q}$	the set rational numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}_+$	the set of nonnegative real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbf{A}$	matrix in $\mathbb{R}^{m\times n}$
$\mathbf{A}_j$	column j of matrix $\mathbf{A}$
$\mathbf{A} \succeq 0$ ($\succ 0$)	$\mathbf{A}$ is positive semidefinite (definite)
x	scalar $x \in \mathbb{R}$
$\mathbf{x}$	vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$
$\boldsymbol{\alpha}$	vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$
$\lvert\boldsymbol{\alpha}\rvert =$	$\sum_{i=1}^n \alpha_i$ for $\boldsymbol{\alpha} \in \mathbb{N}^n$
$\mathbb{N}_d^n \subset \mathbb{N}^n$	the set $\{\boldsymbol{\alpha} \in \mathbb{N}^n : \lvert\boldsymbol{\alpha}\rvert \leq d\}$
$\mathbf{x}^{\boldsymbol{\alpha}}$	vector $\mathbf{x}^{\boldsymbol{\alpha}} = (x_1^{\boldsymbol{\alpha}_1} \cdots x_n^{\boldsymbol{\alpha}_n})$, $\mathbf{x} \in \mathbb{C}^n$ or $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\alpha} \in \mathbb{N}^n$
$\mathbb{R}[x]$	ring of real univariate polynomials
$\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_n]$	ring of real multivariate polynomials
$\Sigma[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$	set of sum-of-squares (SOS) polynomials
$\mathcal{P}(\mathcal{X})_d$	space of polynomials of degree at most d, nonnegative on $\mathcal{X}$
$(\mathbf{x}^{\boldsymbol{\alpha}})$	canonical monomial basis of $\mathbb{R}[\mathbf{x}]$
$V_{\mathbb{C}}(I) \subset \mathbb{C}^n$	the algebraic variety associated with an ideal $I \subset \mathbb{R}[\mathbf{x}]$
$\sqrt{I}$	the radical of an ideal $I \subset \mathbb{R}[\mathbf{x}]$
$\sqrt[\mathbb{R}]{I}$	the real radical of an ideal $I \subset \mathbb{R}[\mathbf{x}]$
$I(V(I)) \subset \mathbb{C}^n$	the vanishing ideal $\{f \in \mathbb{R}[\mathbf{x}] : f(\mathbf{z}) = 0\ \forall \mathbf{z} \in V_{\mathbb{C}}(I)\}$
$V_{\mathbb{R}}(I) \subset \mathbb{R}^n$	the real variety associated with an ideal $I \subset \mathbb{R}[\mathbf{x}]$

$I(V_{\mathbb{R}}(I)) \subset \mathbb{R}[\mathbf{x}]$	the real vanishing ideal $\{f \in \mathbb{R}[\mathbf{x}]: f(\mathbf{x}) = 0, \forall \mathbf{x} \in V_{\mathbb{R}}(I)\}$
$\mathbb{R}_n[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$	real multivariate polynomials of degree at most n
$\Sigma_n(\mathbf{x}] \subset \mathbb{R}_{2n}[\mathbf{x}]$	set of SOS polynomials of degree at most $2n$
$\mathbb{R}[\mathbf{x}]^*$	the vector space of linear forms on $\mathbb{R}[\mathbf{x}]$
$\mathbb{R}_n[\mathbf{x}]^*$	the vector space of linear forms on $\mathbb{R}_n[\mathbf{x}]$
$\mathbf{y} = (y_\alpha) \subset \mathbb{R}$	moment sequence indexed in the canonical basis of $\mathbb{R}[\mathbf{x}]$
$\mathbf{M}_n(\mathbf{y})$	moment matrix of order n associated with the sequence $\mathbf{y}$
$\mathbf{M}_{\mu,n}$	moment matrix of order n associated with a measure μ
$\mathbf{M}_n(g\,\mathbf{y})$	localizing matrix of order n associated with the sequence $\mathbf{y}$ and $g \in \mathbb{R}[\mathbf{x}]$
$P(g) \subset \mathbb{R}[\mathbf{x}]$	preordering generated by the polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$
$Q(g) \subset \mathbb{R}[\mathbf{x}]$	quadratic module generated by the polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$
$\mathcal{M}(\mathbf{K})_n$	space of finite sequences $\mathbf{y} \in \mathbb{R}^{s(n)}$ with a representing measure on $\mathbf{K}$
$\mathcal{P}(\mathbf{K})_n$	space of positive polynomials of degree at most n, nonnegative on $\mathbf{K}$
$\mathscr{B}(\mathbf{X})$ (resp. $\mathscr{B}(\mathbf{X})_+$)	space of bounded (resp. bounded nonnegative) measurable functions on $\mathbf{X}$
$\mathscr{C}(\mathbf{X})$ (resp. $\mathscr{C}(\mathbf{X})_+$)	space of bounded (resp. bounded nonnegative) continuous functions on $\mathbf{X}$
$\mathscr{M}(\mathbf{X})$	vector space of finite signed Borel measures on $\mathbf{X} \subset \mathbb{R}^n$
$\mathscr{M}(\mathbf{X})_+ \subset \mathscr{M}(\mathbf{X})$	space of finite positive Borel measures on $\mathbf{X} \subset \mathbb{R}^n$
$\mathscr{P}(\mathbf{X}) \subset \mathscr{M}(\mathbf{X})_+$	space of Borel probability measures on $\mathbf{X} \subset \mathbb{R}^n$
$L^p(\mathbf{X}, \mu)$	Banach space of functions on $\mathbf{X} \subset \mathbb{R}^n$ such that $(\int_{\mathbf{X}} \|f\|^p d\mu)^{1/p} < \infty$, $1 \le p < \infty$
$L^\infty(\mathbf{X}, \mu)$	Banach space of measurable functions on $\mathbf{X} \subset \mathbb{R}^n$ such that $\\|f\\|_\infty := \text{ess sup } \|f\| < \infty$
$\sigma(\mathcal{X}, \mathcal{Y})$	weak topology on $\mathcal{X}$ for a dual pair $(\mathcal{X}, \mathcal{Y})$ of vector spaces
$\nu \ll \mu$	ν is absolutely continuous with respect to μ (for measures)
$\nu \perp \mu$	measures ν and μ are mutually singular

Preface

Among the many positive-definite kernels appearing in classical analysis, approximation theory, probability, mathematical physics, control theory and more recently in machine learning, the Christoffel–Darboux (CD) kernel stands out by its numerical accessibility from raw data and its versality in encoding/decoding fine properties of the generating measure. Indeed this unique feature was recognized very early and was exploited over a century and a half via surprising developments. One can safely draw a parallel: just as the power moment problem is the quintessential inverse problem, the CD kernel is the prototypical positive-definite reproducing kernel. While the computationally oriented practitioner may think that dealing with real monomials brings instability, we argue that complex monomials restricted to the unit circle or higher-dimensional tori are the very stable ubiquitous Fourier modes.

The CD kernel has a particular property that enables us to identify the underlying reproducing kernel Hilbert space (RKHS) inner product with a bilinear form induced by a given measure over a finite-dimensional function space. This feature allows us to develop a rich theory describing the relation between the Christoffel–Darboux kernel and the underlying measure in a data analysis context. Our aim in this book is to explain this property and its application in data analysis and the numerical treatment of statistical data. Another of our goals is to make it straightforward for the non-expert reader to obtain further insights about the role of the Christoffel function in function theory, approximation theory and the spectral analysis of dynamical systems, as well as sketching some possible extensions (e.g. to Lebesgue $L^p(\mu)$ spaces).

Since the Christoffel function and the Christoffel–Darboux kernel have a long mathematical history, we confine ourselves on the one hand to describing in a streamlined manner their classical theory and on the other to giving an up-to-date collection of results that provide a theoretical basis for such applications. Examples of these include:

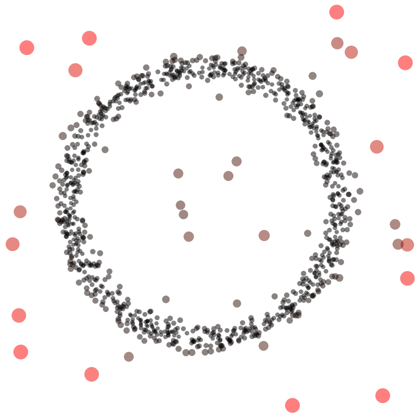


Figure 1 $N = 1040$ and $n = 8$; points with size and color proportional to the value of $1/\Lambda_n^{\mu_N}$.

- *outlier detection*, where $\Lambda_n^{\mu_N}$ provides a simple test to decide whether a point $\mathbf{x}$ of the cloud can be considered as an outlier;
- *density estimation*, when μ_N is the empirical measure μ_N on a cloud of points drawn for some unknown probability distribution μ on $\mathbf{\Omega}$, with a density with respect to Lebesgue measure on $\mathbf{\Omega}$;
- *manifold learning*: when the cloud of points is supported on a subset of a manifold (e.g. the sphere) or on an algebraic variety, can we detect the manifold (or algebraic variety) and its dimension?

To better appreciate the simplicity of the approach, consider the cloud of two-dimensional points shown in Figure 1. Most points are in an annulus and the color and size of a point $\boldsymbol{\xi}$ is proportional to the value of $\Lambda_n^{\mu_N}(\boldsymbol{\xi})^{-1}$ at this point. Therefore all points $\boldsymbol{\xi}$ with "color" $\Lambda_n^{\mu_N}(\boldsymbol{\xi}) \leq \tau$ (for some threshold τ) are declared potential *outliers*. In Figure 1 one clearly sees that points with colors close to pink, red, or brown, are "outside" the annulus.

Of course, when μ_N is the empirical measure supported on a sample drawn from some distribution μ on $\mathbf{\Omega}$, to obtain rigorous asymptotic results on the unknown μ and $\mathbf{\Omega}$, it is expected that one has to carefully scale the degree n with the size N of the sample. This issue is clearly particular to data analysis because one uses an empirical measure μ^N on a typically *finite* sample. We show how such asymptotic results can be rigorously justified in this data analysis framework.

By its nature, our text interlaces distant themes, over-simplifies most of the theoretical background and sacrifices fine tuning for wide accessibility and utility. We are aware that balancing such opposite tendencies leads to brutal omissions. The story does not end here. We apologize in advance to the neglected parties and invite them to take our essay as a basis for exploring novel ramifications of Christoffel–Darboux analysis.

Having said all that, we have to recognize the lasting creative power of the two founders of the theory. The genius of E. B. Christoffel emanates from every page of the astounding collection Butzer and Fehér (1981). Darboux's innovative brilliance was recognized by all leading figures of the mathematical landscape (Lippmann et al., 1912). His eulogy in Hilbert (1920) is as fresh and accurate now as it was at the time of its publication a century ago.

Acknowledgements: The first and second authors gratefully acknowledge financial support from **ANITI** (Artifical & Natural Intelligence Toulouse Institute), coordinated by the Federal University of Toulouse within the framework of the French Program "Investing for the Future C PIA3" under the grant agreement ANR-19-PIA3-0004. The third author was partially supported by a Simons Foundation collaboration grant for mathematicians.

1
Introduction

To get a glimpse of the main theme of the book, consider an arbitrary cloud of N points $\mathbf{x}_i = (x_i, y_i) \in \mathbb{R}^2$, $i = 1, \ldots, N$, in the plane, dense enough to form some geometric shape. For instance in Figure 1.1 the shape looks like a rotated letter "T"; similarly in the frontispiece, the cloud of points is concentrated on the letters "C" and "D" (for Christoffel and Darboux). Then we invite the reader to perform the following simple operations on the preferred cloud of points:

1. Fix $n \in \mathbb{N}$ (for instance $n = 2$) and let $s(n) = \binom{n+2}{2}$.
2. Let $\mathbf{v}_n(\mathbf{x}) = (1, x, y, x^2, x\,y, \ldots, x\,y^{n-1}, y^n)$ be the vector of all monomials $x^i y^j$ of total degree $i + j \leq n$.
3. Form $\mathbf{X}_n \in \mathbb{R}^{n \times s(n)}$, the design matrix whose ith row is $\mathbf{v}(\mathbf{x}_i)$, and the real symmetric matrix $\mathbf{M}_n \in \mathbb{R}^{s(n) \times s(n)}$ with rows and columns indexed by monomials such that
$$\mathbf{M}_n := \frac{1}{N} \mathbf{X}_n^T \mathbf{X}_n.$$
4. Form the polynomial
$$\mathbf{x} \mapsto p_n(\mathbf{x}) := \mathbf{v}_n(\mathbf{x})^T \mathbf{M}_n^{-1} \, \mathbf{v}_n(\mathbf{x}).$$
5. Plot the level sets $S_\gamma := \{\mathbf{x} \in \mathbb{R}^2 : \; p_n(\mathbf{x}) = \gamma\}$ for some values of γ, and in red for the particular value $\gamma = \binom{2+n}{2}$.

As the reader can observe in Figure 1.1, the various level sets (and in particular the red one) capture quite accurately the shape of the cloud of points.

The above polynomial p_n is associated with the cloud of points $(\mathbf{x}_i)_{i \leq N}$ only via the real symmetric matrix $\mathbf{M}_n$ in a conceptually simple manner, the main