Philipps-Universität Marburg
Fb Mathematik u. Informatik

Nonparametric estimation in models for unobservable heterogeneity

Kumulative Dissertation

zur Erlangung des Doktorgrades der Mathematisch-
Naturwissenschaftlichen Fakultäten der
Philipps-Universität Marburg

vorgelegt von

Daniel Hohmann

aus Hünfeld

Marburg, Dezember 2013

Erstgutachter: Prof. Dr. Hajo Holzmann
Zweitgutachter: Prof. Dr. Melanie Birke

Tag der Disputation: 28. Februar 2014

Acknowledgments

First of all I would like to thank Prof. Dr. Hajo Holzmann for supervising this thesis, for numerous most helpful discussions, and for having provided the opportunity to participate in several conferences and workshops. Also, I would like to thank Prof. Dr. Melanie Birke for making the second assessment of this thesis. Special thanks, of course, go to all my current and former colleagues, especially to Grigory Alexandrovich, Matthias Eulert, Tobias Filusch, Anna Leister, Max Nattermann, Hilke Stibbe, Dirk Engel, Dr. Florian Schwaiger, and Dr. Florian Ketterer. Thank you for the most pleasant working atmosphere, the coffee and ping-pong breaks, but certainly also for the numberless technical exchanges. Thanks to Thorsten Ante and Johannes Berens for reading parts of this thesis and giving valuable comments. And last but not least I would like to thank Kristin Kunze, Annika Hohmann, and my parents Elisabeth and Paul Hohmann for supporting me in many respects.

Thank you.

Contents

Part I

Introduction

1

Nonparametric models and unobservable heterogeneity

Statistical models are frequently used to describe the behavior of or interdependencies between individuals, e. g. economic agents. In order to make statistical inference in such models more tractable, it is most often assumed that the observed individuals act almost identically so that, exemplarily in a regression model, their behavior can be expressed in terms of a deterministic link function. This restriction might be crucial in practice, though. As for instance already recognized by Klein (1953), naturally present differences between individuals in a cross-section sample can hardly be explained by a simple regression including a manageable size of covariates, particularly not by means of fixed effects, see also Section 1.2 below. Hence, in some situations it may be reasonable to proceed to the application of statistical models which account for unobservable heterogeneity within the observed population.

1.1. Finite mixtures

Finite mixture models certainly provide the most natural way to describe populations with unobservable heterogeneity. Assume that the distribution function of a random variable (or vector) Y is given by the finite sum

$$F(y) = \sum_{i=1}^{k} \pi_i F_i(y), \tag{1.1}$$

where $\pi_i \geq 0$, $\sum_{i=1}^{k} \pi_i = 1$, and each $F_i : \mathbb{R} \to [0,1]$ being a distribution function itself. Model (1.1) means that the observable population can be divided into k latent sub-populations, the F_i correspond to the distributions within these sub-populations, the weights π_i determine their percentage shares on the overall population. The problem now is to draw inference about the distribution functions F_i and the weights π_i, and conceivably the number of components k if unknown, from a finite sample of observations of Y.

Applications of mixture models are widespread, including cluster and latent class analyses, discriminant analysis, as well as image and survival analyses, successfully used in economics, biology, astronomy, medicine, genetics, and many more. In practice, the F_i are most often supposed to belong to some parametric family, the location-scale family of normal distributions being the most prominent candidate, which turns the estimation in model (1.1) into a fully parametric problem. Estimates can then be obtained via maximum likelihood, minimum chi-squared, or moment methods. There is extensive literature on this parametric topic, see for instance the monographs of Everitt and Hand (1981), Titterington, Smith and Makov (1985), or McLachlan and Peel (2000).

The choice of an adequate parametric family for the F_i may be crucial, though, especially if there is no further information on the sub-populations available. This might be one reason why in recent years the interest in mixtures with nonparametric components has grown significantly. In general, however, model (1.1) is not nonparametrically identifiable, i. e., without further shape constraints imposed on the component distributions F_i, the representation of F in (1.1) is not unique. This in turn renders the construction of consistent estimators impossible as long as there is no training data available for all the sub-populations (Murray and Titterington, 1978; Hall, 1981).

Not assuming training data to be on hand, Hall and Zhou (2003) are the first to study estimation in (1.1) without parametric assumptions on the F_i. They particularly consider a multivariate mixture with two components, each component having independent coordinates, which they propose for modeling repeated medical tests on a person with unknown disease status. Under mild regularity conditions, they show identifiability in case that the observations' dimension is at least three, and provide semiparametric, root-n-consistent estimators. Bordes, Mottelet and Vandekerkhove (2006) as well as Hunter, Wang and Hettmansperger (2007) consider the case of univariate mixtures, and in order to obtain identifiability they assume that all components belong to the same location family of one zero symmetric distribution. A closely related two-component mixture is considered by Bordes and Vandekerkhove (2010) and Hohmann and Holzmann (2013a), who no longer assume that the symmetric components are taken from the same location family, but that one of the component distributions is known in advance, possibly up to a location parameter. Consistent estimators are constructed for all these univariate, semiparametric models.

The assumption of symmetric components, which all the latter models for univariate data have in common, may be crucial in several fields of application, e. g. finance, though. Henry, Kitamura and Salanié (2010) and later Hohmann and Holzmann (2013b, cf. Chapter 3) therefore follow a different approach. They consider a two-component conditional mixture model, where the additional information provided by the observation of covariates is used to obtain identifiability. In particular, the basic model assumption is that the distributions within the sub-populations do not vary with the covariates, only the mixture proportions $\pi_i = \pi_i(z)$ do. This assumption combined with suitable tail dominance shape constraints imposed on the F_i successfully identifies all the model parameters and helps to construct fully nonparametric estimators. The shape constraints considered are typically satisfied by location-scale-type mixtures of supersmooth distributions, exemplarily including the skew-normal distributions.

1.2. Random coefficient regression

The model most popularly used to explain dependencies between features of individuals is the linear regression model. Let Y and X be observable with values in $\mathbb{R}$ and $\mathbb{R}^d$, respectively, and assume that the conditional mean of Y given X is a linear function in X,

$$Y = \beta' X + \varepsilon = \beta_1 X_1 + \ldots + \beta_d X_d + \varepsilon, \tag{1.2}$$

where β is an unknown vector of effects that is to be estimated, and ε is an additive noise having zero conditional mean, $\mathrm{E}(\varepsilon|X) = 0$. The X_i are referred to as explanatory variables or covariates, Y as response variable. The error ε is assumed to comprise measurement errors as well as any other exogenous variables affecting Y which are not correlated with X. Given a finite sample of copies of (Y,X), the factors β_i in (1.2) can then be estimated via least-squares, maximum-likelihood, or related techniques. Regression is usually applied for descriptive purposes or prediction, in fields such as socioeconomics, finance, ecology, neuroscience, and epidemiology, among many others. For a comprehensive review on regression models and its applications see for instance the very recent monograph of Fahrmeir, Kneib, Lang and Marx (2013).

In point of fact, the individuals which observations are drawn from are most often likely to be heterogeneous. Furthermore, it may be unrealistic to assume that X is indeed uncorrelated with all excluded influence quantities. Thus, the coefficients β_i ought to be allowed to vary across the individuals. Introducing a separate vector of effects for each observation is inconvenient due to the increasing number of parameters, though. The attractive midway therefore is to treat the β_i as being random, and then to draw inference on parameters of their distributions. This is called the random coefficient model, which in its simplest form can be written as

$$Y = \beta' X. \tag{1.3}$$

Note that it is no longer necessary to attach the error ε as in (1.2) since it can be incorporated into (1.3) as a random intercept by setting $X_1 = 1$.

Random coefficient regression was already studied by Hildreth and Houck (1968) and Swamy (1970, 1971). Basically, under the assumptions of mean independence and homoscedasticity, i. e. $\mathrm{E}(\beta|X) = \mu_\beta$ and $\mathrm{Cov}(\beta|X) = \Sigma_\beta$, they identify mean and variance of each effect β_i, and provide consistent estimates. Ramanathan and Rajarshi (1992) for example relax the assumption of homoscedasticity. Estimating other features of the distribution of β are still not taken into account, though.

Only recently, assuming the effects β to be continuously distributed and independent of X, Hoderlein, Klemelä and Mammen (2010) propose a method for nonparametrically estimating the joint density f_β of the random effects β_i. Their approach is based on the observation that the Radon transform $\mathrm{R}f_\beta$ of the density f_β, describing the integrals of f_β along arbitrary vector hyperplanes, is given by the conditional density of (adequately transformed versions of) Y given X. Since it is well known that a function on $\mathbb{R}^d$ is uniquely determined by all its integrals along hyperplanes (Radon, 1917), this helps to identify f_β in (1.3), and then to construct a sample counterpart estimator $\hat{f}_\beta$ using a regularized inverse of the operator R. Their estimator is shown to be rate optimal in certain Sobolev spaces, and its asymptotic normality is provided.

As the authors bring up themselves, the estimator constructed in Hoderlein et al. (2010) has a significant drawback, though. For simplicity, let X assume values in $\mathbb{R}^2$, and particularly include an intercept into (1.3), i. e. $X_1 = 1$. In this simple case, the asymptotic results for $\hat{f}_\beta$ do only hold true if X_2 is heavy-tailed, e. g. cauchy-type tails, which seems to be quite restrictive. As argued in Hohmann and Holzmann (2013d), this is due to the fact that the density f_Φ of the angle Φ defined by $\sin(\Phi) = X_2/\|X\|$ is no longer bounded away from zero if X_2 does not exhibit heavy tails. They study the Radon transform as operator between suitably weighted L_2-spaces, and thereby give an insight into how the degree of ill-posedness for the inversion problem of estimating f_β in fact strongly depends on the behavior of f_Φ close to zero as they provide corresponding minimax rates on certain Sobolev spaces in a related white noise model.

1.3. Organization of the thesis

This cumulative doctoral thesis is organized as follows. Short summaries of the two publications involved can be found in the subsequent Chapter 2. Both summaries are given in English and German. The principal cumulative part is Part II, comprising the publications 'Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation', published in the Electronic Journal of Statistics (Hohmann and Holzmann, 2013b), and 'Weighted angle Radon transform: Convergence rates and efficient estimation', submitted (Hohmann and Holzmann, 2013d). With the exception of numerations and minor corrections, both papers are presented in their original version as published resp. submitted. Part III contains supplementary material for the first of the two publications which is not part of the published version but can be downloaded from the web (Hohmann and Holzmann, 2013c).

2

Summary of publications

2.1. Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation

English. Assume that the distribution of a random variable Y conditional on a given vector Z of covariates can be expressed as the two-component mixture

$$F(y|z) = (1-\pi(z))F_0(y) + \pi(z)F_1(y), \tag{2.1}$$

where the distributions F_0 and F_1 within the two latent sub-populations are not allowed to depend on the realization of Z. This model was recently studied by Henry, Kitamura and Salanié (2010) who established nonparametric identifiability of F_0, F_1, and π in (2.1) for Z being a binary regressor together with tail dominance assumptions on the component distributions F_0 and F_1 which typically apply to location-type mixtures. Further, nonparametric, pointwise asymptotically normal estimators were provided.

In Hohmann and Holzmann (2013b), see Chapter 3, we extend the results of Henry et al. (2010) as follows. First, we drop the assumptions on Z which may now be arbitrarily distributed. Second, beside those imposed in Henry et al. (2010), we also consider tail dominance assumptions which involve the Fourier transforms $\tilde{F}_0$ and $\tilde{F}_1$ of F_0 and F_1, resp., and which are naturally satisfied by more general location-scale-type mixtures. We show identifiability of F_0, F_1, and π in (2.1) under these weaker assumptions, and construct fully nonparametric estimates based on a sample of independent copies of (Y,Z). Depending on tuning parameters that in particular determine the rates of convergence, asymptotic normality of these estimators is established. In fact, the centered and adequately rescaled empirical processes $\sqrt{r_{i,n}}(\hat{F}_i - F_i)$ are shown to converge to a weak normal limit in $\ell^\infty(\mathbb{R})$, which might be of special interest for the construction of confidence bands, for example.

Our main motivation for studying (2.1) is the multivariate two-component mixture model with independent coordinates as considered by Hall and Zhou (2003), which they introduced to model measurements of repeated tests on a person with unknown disease status. They proved nonparametric identifiability of their model under mild assumptions in case that the dimension of the observations is at least three, only partial identifiability results were available in two dimensions. We show how their multivariate model can be cast into the framework of the conditional mixture (2.1), and how this can be utilized in order to identify and estimate the mixture weight as well as the component distributions in any coordinate, even in two dimensions, provided our tail assumptions hold.

The upcoming difficulty when analyzing the asymptotics of nonparametric characteristic function based estimators is the treatment of the corresponding empirical characteristic processes, which is significantly more involved than that of ordinary empirical distribution functions. In fact, our main theoretical contribution is the large sample theory for estimates of a limit of quotients of two characteristic functions in the tails, say $\lim_{y\to\infty} \tilde{G}(y)/\tilde{H}(y)$, based on mutually independent i.i.d. samples drawn from G and H. For this we use strong approximation techniques in $\ell^\infty(\mathbb{R})$, involving certain complex Gaussian processes, and give an L_2-entropy bound on the corresponding class of generating functions, which allows to evaluate these latter processes at random levels that tend to infinity as the number of observations increases. Another important technical ingredient for the derivation of joint weak limits is the notion of asymptotic independence of random sequences.

To conclude, a simulation study is conducted in order to illustrate the theoretical large sample results, which we do in the multivariate model of Hall and Zhou (2003). It is particularly pointed out how the choice of the tuning parameters strongly affects the performance of the estimators. As this dependence is crucial in practice where no information on the true component distributions is available, though, we devise a repeated random sub-sampling cross-validation scheme which allows to choose these parameters in a data-driven way. While there is no theoretical analysis for this approach in our model on hand, the simulation results seem to give a numerical justification.

German. Wir nehmen an, dass die bedingte Verteilung einer Zufallsvariable Y gegeben einem Vektor Z von Kovariablen durch die Zwei-Komponenten-Mischung

$$F(y|z) = (1-\pi(z))F_0(y) + \pi(z)F_1(y) \tag{2.2}$$

beschrieben werden kann, wobei insbesondere vorausgesetzt wird, dass die Verteilungsfunktionen F_0 und F_1 innerhalb der latenten Subpopulationen nicht von der tatsächlichen Realisierung von Z abhängen. Dieses Modell wurde kürzlich von Henry, Kitamura and Salanié (2010) untersucht. Unter der Annahme, dass Z ein binärer Regressor ist, und dass die Komponenten F_0 und F_1 gewisse Dominanzen in ihren Schwänzen erfüllen, welche überlicherweise auf Lokationsmischungen zutreffen, wurden F_0, F_1 und π in (2.2) identifiziert und entsprechende nichtparametrische, punktweise asymptotisch normale Schätzer konstruiert.

In Hohmann and Holzmann (2013b), siehe Chapter 3, erweitern wir die Resultate von Henry et al. (2010) wie folgt: Zum einen verzichten wir auf jegliche strukturelle Voraussetzungen an die Kovariablen Z. Zum anderen betrachten wir nicht nur Annahmen an das Dominanzverhalten der Verteilungsfunktionen F_0 und F_1, sondern auch an das deren Fouriertransformierten $\hat{F}_0$ und $\hat{F}_1$, welche typischerweise auf allgemeinere Lokations-Skalenmischungen zugeschnitten sind. Wir identifizieren F_0, F_1 und π in (2.2) unter diesen schwächeren Annahmen und konstruieren auch hier entsprechende Schätzer. In Abhängigkeit von Regulierungsparametern, die unter anderem Einfluss auf die Konvergenzgeschwindigkeit haben, zeigen wir die asymptotische Normalität unserer Schätzer, und zwar speziell, dass die zentrierten und passend skalierten empirischen Prozesse $\sqrt{r_{i,n}}(\hat{F}_i - F_i)$ einen schwachen, normalverteilten Limes in $\ell^\infty(\mathbb{R})$ besitzen, was beispielsweise für die Konstruktion von gleichmäßigen Konfidenzbändern verwandt werden könnte.

Unsere Hauptmotivation für die Betrachtung von (2.2) ist das multivariate Zwei-Komponenten-Mischungsmodell mit unabhängigen Koordinaten, wie es Hall and Zhou (2003) für die Modellierung der Messergebnisse wiederholter Tests an einem Patienten mit unbekanntem Gesundheitszustand einführen. Sie zeigen die Identifizierbarkeit ihres Modells unter schwachen Annahmen für den Fall, dass die Dimension der beobachteten Daten mindestens drei ist. Für den zweidimensionalen Fall erhalten sie partielle Identifizierbarkeit. Wir zeigen, wie sich dieses multivariate Modell in den Kontext von bedingten Mischungen, insbesondere (2.2), einbinden lässt und wie dies dazu verwandt werden kann, um die Verteilungsfunktionen aller Koordinaten beider Komponenten sowie das Mischgewicht zu schätzen, sogar in zwei Dimensionen, vorausgesetzt unsere Annahmen an die Verteilungsschwänze der Mischungskomponenten sind erfüllt.

Die Schwierigkeit, die bei der Untersuchung des asymptotischen Verhaltens von Schätzern, die auf der nichtparametrischen Schätzung von charakteristischen Funktionen basieren, auftritt, ist die Handhabung der entsprechenden empirischen charakteristischen Prozesse, die bedeutend schwieriger ist als die von gewöhnlichen empirischen Verteilungsfunktionen. In diesem Zusammenhang ist unser theoretischer Hauptbeitrag die asymptotische Theorie für das Schätzen des Grenzwertes eines Quotienten von zwei charakteristischen Funktionen in deren Schwänzen, sagen wir $\lim_{y\to\infty}\tilde{G}(y)/\tilde{H}(y)$, basierend auf zwei gegenseitig unabhängigen Folgen von unabhängigen und identisch nach G bzw. H verteilten Beobachtungen. Hierzu verwenden wir starke Approximationen in $\ell^\infty(\mathbb{R})$ durch entsprechende, komplexwertige Gaußprozesse und bestimmen eine L_2-Informationsschranke für die Klasse der prozessgenerierenden Funktion zur Auswertung dieser Gaußprozesse an zufälligen Stellen, die von den Beobachtungen abhängen und mit wachsender Anzahl an Beobachtungen gegen unendlich gehen. Technisch interessant ist außerdem das Konzept von asymptotisch unabhängigen Folgen von Zufallsvariablen, was die Herleitung von gemeinsamen schwachen Limiten vereinfacht oder erst ermöglicht.

In dem Modell von Hall and Zhou (2003) haben wir abschließend eine Simulationsstudie zur Veranschaulichung der theoretischen Ergebnisse durchgeführt. Besonders gut zu erkennen ist dabei der Einfluss der Regulierungsparameter auf die Güte der Schätzer. Da diese Abhängigkeit in der Praxis, wo keine Informationen über die wahren Verteilungen bekannt sind, kritisch ist, haben wir außerdem ein Kreuzvalidierungsschema entwickelt, was es ermöglicht, die entsprechenden Parameter datengetrieben zu wählen. Obwohl es zu diesem Verfahren in unserem Modell noch keine theoretische Untersuchung gibt, scheinen die Simulationsergebnisse eine numerische Rechtfertigung zu liefern.

2.2. Weighted angle Radon transform: Convergence rates and efficient estimation

English. Given an integrable function f on the unit disc $B_1(0)$ in $\mathbb{R}^2$, the Radon transform $\mathrm{R}f$ of f is defined as

$$\mathrm{R}f : [-\pi/2,\pi/2]\times[-1,1]\to\mathbb{R},$$
$$(\varphi,s)\mapsto\int_{|t|\le\sqrt{1-s^2}} f(s\cos\varphi - t\sin\varphi, s\sin\varphi + t\cos\varphi)\,dt,$$

i. e. $(\mathrm{R}f)(\varphi,s)$ is the line integral of f along the hyperplane of those $(x,y)\in B_1(0)$ for which $x\cos\varphi + y\sin\varphi = s$. Recovering a function f from observations of its Radon transform is an issue in quite a number of applications, including X-ray transmission computed and emission computed tomography, astronomy, optics, and geophysics. See Deans (1983) and Natterer (1986) for a detailed review.

Inverting the operator R is usually considered to be a very mildly ill-posed inverse problem. In Hohmann and Holzmann (2013d), however, see Chapter 4, we argue that there are several statistical models involving the Radon transform, e. g. regression models for computerized tomography, nonparametric random coefficient regression models, and density estimation models for positron emission tomography, where the measurement design for obtaining the observations may naturally lead to much more ill-posed even up to severely ill-posed inverse problems. For this, we study R as an operator between suitably weighted L_2-spaces,

$$\mathrm{R} : L_2(B_1(0);\mu_2) \longrightarrow L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1), \tag{2.3}$$

where the measure $d\mu_1(\varphi,s) = \lambda(\varphi)w_1^\gamma(s)\,d\varphi\,ds$ is stipulated by the measurement design, and $d\mu_2(x,y) = w_2^\gamma(x,y)\,dx\,dy$ is then chosen for technical reasons in order to make the singular value decomposition of R tractable. Stemming from the framework of computerized tomography, commonly $L_2(B_1(0);\mu_2)$ is called *brain space* and $L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1)$ *detector space*. The weight functions w_1^γ and w_2^γ are specific, parametric families in $\gamma > -1/2$ which we adopted from earlier works, cf. Davison (1981). The most striking feature, however, hitherto hardly treated in the literature, is the almost arbitrary weight function λ on the angle φ.

Depending on γ and the weight λ, we provide the singular value decomposition of the operator R in (2.3). It turns out that the singular values are closely related to the eigenvalues of the sequence of Toeplitz matrices generated by λ. In particular, if the weight λ is not bounded away from zero, then these eigenvalues are not, either, and as a consequence the rate of ill-posedness for inverting R increases. Based on well known facts as well as only recently obtained results for the asymptotics of the eigenvalues of sequences of Toeplitz matrices, we derive bounds on the decay rate of the singular values of R for certain classes of weight functions in order to gain an insight into the actual degree of ill-posedness. We particularly find that if λ only has isolated zeros, typically at the boundary points $\pm\pi/2$, then the inverse problem remains mildly ill-posed, and the degree of ill-posedness is basically determined by the degree of the roots. If λ has bounded support, though, as it is for example the case for the so called limited angle Radon transform (cf. Davison, 1983), then the singular values decay exponentially fast so that the inverse problem becomes severely ill-posed. We prove that the operator R in (2.3) is injective for any $\gamma > -1/2$ whenever the weight function λ is not essentially zero, and therewith derive explicit formulas for the singular functions in both detector and brain space, which help to define ellipsoid type smoothness conditions for functions in brain space, and which are important from a practical point of view in order to construct SVD-based estimates.

We then consider estimating a function f from a noisy observation Y of Rf in the Gaussian white noise model

$$dY(\varphi,s) = (\mathrm{R}f)(\varphi,s)\,d\mu_1(\varphi,s) + \varepsilon\,dW(\varphi,s), \tag{2.4}$$

where W is a weighted Brownian sheet on $[-\pi/2,\pi/2]\times[-1,1]$ such that, for g_1 and g_2 in detector space, the variables $\int g_1\,dW$ and $\int g_2\,dW$ are jointly Gaussian with zero mean and covariance $\langle g_1,g_2\rangle_{\mu_1}$, and where $\varepsilon > 0$ is the noise level. Model (2.4) means that, for any function g in detector space, we may observe the random variable $\int g\,dY$. In particular, plugging in the complete system of orthonormal singular functions in detector space, we obtain a sequence of noisy observations of the Fourier coefficients of f with respect to the system of singular base functions in brain space, the noise being i.i.d. Gaussian.

While this white noise model, in which theoretical analyses are less complicated, is only an idealized large-sample approximation to standard models for nonparametric regression or density estimation based on indirect observations, it still gives a valuable insight into the difficulty of the estimation problem. Based on the asymptotic behavior of the singular values of R, we provide upper and lower bounds on the minimax rates for estimating f in certain Sobolev-type smoothness classes under mild assumptions on the weight function λ in case of isolated zeros at $\pm\pi/2$, for any $\gamma > -1/2$. Although exact minimax rates and efficiency constants can only be given for $\gamma = 1$ and by imposing much stronger conditions, e. g. assuming λ to be banded and the root being of order two, in any of these mildly ill-posed cases, though, we find that the well known Pinsker estimator remains asymptotically efficient. Finally, in the severely ill-posed case when λ has bounded support, we even provide the exact minimax rates for any choice of $\gamma > -1/2$, and for the special case $\lambda = \mathbf{1}_{[-\eta,\eta]}$ for some $\eta < \pi/2$ (limited angle case) we show that a simple projection estimator is not only efficient but even adaptive in the exact minimax sense on the smoothness classes considered.

German. Die Radontransformation einer auf der Einheitsscheibe $B_1(0)$ im $\mathbb{R}^2$ integrierbaren Funktion f ist definiert als

$$\begin{aligned} \mathrm{R}f &: [-\pi/2,\pi/2]\times[-1,1]\to\mathbb{R}, \\ (\varphi,s) &\mapsto \int_{|t|\le\sqrt{1-s^2}} f(s\cos\varphi - t\sin\varphi, s\sin\varphi + t\cos\varphi)\,dt, \end{aligned}$$

d. h. $(\mathrm{R}f)(\varphi,s)$ ist das Linienintegral entlang der Hyperebene all derer $(x,y)\in B_1(0)$, für die $x\cos\varphi + y\sin\varphi = s$ gilt. Für viele Anwendungen ist es erforderlich, eine Funktion f aus der Beobachtung ihrer Radontransformierten zu rekonstruieren, unter anderem in der Computertomografie, Astronomie, Optik und Geophysik. Einen detaillierten Überblick findet man beispielsweise in Deans (1983) und Natterer (1986).

Das Invertieren des Operators R wird gewöhnlich als ein sehr schwach schlecht gestelltes inverses Problem angesehen. In Hohmann and Holzmann (2013d), siehe Chapter 4, stellen wir jedoch heraus, dass es eine Vielzahl an statistischen Modellen gibt, zu denen insbesondere Regressionsmodelle für die Computertomografie, nichtparametrische Regressionsmodelle mit zufälligen Koeffizienten sowie Modelle für die Dichteschätzung in der Positronen-Emissions-Tomographie zählen, bei denen das Design der Messung zur Gewinnung der Beobachtungen auf natürliche Weise zu einem deutlich schlechter gestellten bis hin zu einem ernsthaft schlecht gestellten inversen Problem führen kann. Hierfür betrachten wir R als Operator zwischen passend gewichteten L_2-Räumen,

$$\mathrm{R}: L_2(B_1(0);\mu_2) \longrightarrow L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1), \tag{2.5}$$

wobei das Maß $d\mu_1(\varphi,s) = \lambda(\varphi)w_1^\gamma(s)\,d\varphi\,ds$ durch das Design der Messung festgelegt ist und die Wahl $d\mu_2(x,y) = w_2^\gamma(x,y)\,dx\,dy$ dann aus technischen Gründen erfolgt, um die Singulärwertzerlegung von R zugänglich zu machen. Bezug nehmend auf den Bereich Computertomografie wird der Raum $L_2(B_1(0);\mu_2)$ gewöhnlich als *brain space* und $L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1)$ als *detector space* bezeichnet. Die Gewichtsfunktionen w_1^γ und w_2^γ sind spezielle, parametrische Familien in $\gamma > -1/2$ und wurden von früheren Arbeiten übernommen, vgl. Davison (1981). Die jedoch interessanteste Zutat in (2.5), die so in der Literatur noch kaum behandelt wurde, ist die nahezu beliebige Gewichtsfunktion λ auf dem Winkel φ.

In Abhängigkeit von γ und der Gewichtsfunktion λ berechnen wir die Singulärwertzerlegung des Operators R in (2.5). Wie sich herausstellt, stehen die Singulärwerte in engem Zusammenhang mit den Eigenwerten der Folge von Toeplitz-Matrizen, die durch λ erzeugt werden. Insbesondere sind diese nicht von Null wegbeschränkt, wenn auch λ nicht von Null wegbeschränkt ist, so dass sich in diesem Fall der Grad der Schlechtgestelltheit für die Invertierung von R weiter erhöht. Basierend auf sowohl gut bekannten wie auch aktuellen Resultaten zum Konvergenzverhalten der Eigenwerte von Folgen von Toeplitz-Matrizen leiten wir Schranken für das Abklingverhalten der Singulärwerte von R für verschiedene Klassen von Gewichtsfunktionen λ her. Diese geben wiederum einen Aufschluss über den tatsächlichen Grad der Schlechtgestelltheit des inversen Problems. Insbesondere stellt sich heraus, dass das Problem schwach schlecht gestellt bleibt, so lange λ nur isolierte Nullstellen, typischerweise an den Randpunkten $\pm\pi/2$, besitzt, wobei der Grad der Schlechtgestelltheit in dem Fall von dem Grad der Nullstellen abhängt. Hat λ jedoch einen beschränkten Träger, so fallen die Singulärwerte exponentiell schnell und das inverse Problem wird ernsthaft schlecht gestellt. Wir zeigen, dass der Operator R in (2.5) für beliebiges $\gamma > -1/2$ injektiv ist, vorausgesetzt die Gewichtsfunktion λ ist nicht essentiell Null, und leiten damit explizite Formeln für die Singulärfunktionen in brain und dector space her. Diese sind insbesondere wichtig für die Konstruktion von SVD-basierten Schätzern.

Anschließend betrachten wir das Schätzen einer Funktion f aus einer verrauschten Beobachtung Y deren Radontransformierten $\mathrm{R}f$ in dem Gaussian-white-noise-Modell

$$dY(\varphi,s) = (\mathrm{R}f)(\varphi,s)\,d\mu_1(\varphi,s) + \varepsilon\,dW(\varphi,s), \tag{2.6}$$

wobei W ein gewichtetes Brown'sches Blatt auf $[-\pi/2,\pi/2]\times[-1,1]$ ist, so dass für beliebige Funktionen g_1 und g_2 im detector space die Zufallsvariablen $\int g_1\,dW$ und $\int g_2\,dW$ gemeinsam normalverteilt sind mit Erwartungswert null und Kovarianz $\langle g_1, g_2\rangle_{\mu_1}$. $\varepsilon > 0$ wird als Geräuschpegel bezeichnet. Das Modell (2.6) besagt insbesondere, dass für jede Funktion g im detector space die Zufallsvariable $\int g\,dY$ beobachtet werden kann, so dass man durch Einsetzen des vollständigen Systems orthonormaler Singulärfunktionen im detector space eine Folge i.i.d. normalverteilt verrauschter Beobachtungen der Fourierkoeffizienten von f bezüglich der Basis von Singulärfunktionen im brain space erhält.

Während dieses von theoretischer Seite her leichter zugängliche white-noise-Modell nur eine Approximation an Standardmodelle für nichtparametrische Regression oder Dichteschätzung basierend auf indirekten Beobachtungen darstellt, gibt es sehr wohl einen wertvollen Einblick in die Schwierigkeit des zugrundeliegenden Schätzproblems. Basierend auf dem asymptotischen Verhalten der Singulärwerte von R bestimmen wir obere und untere Schranken für die Minimax-Raten zum Schätzen von f in diversen Sobolev-Glattheitsklassen unter schwachen Annahmen an die Gewichtsfunktion λ im Falle von isolierten Nullstellen bei $\pm\pi/2$ für beliebiges $\gamma > -1/2$. Obwohl exakte Raten und Effizienzkonstanten nur für den Fall $\gamma = 1$ und unter deutlich stärkeren Annahmen an λ berechnet werden können, stellt sich heraus, dass der bekannte Pinsker-Schätzer in all diesen schwach schlecht gestellten Problemen asymptotisch effizient bleibt. Abschließend betrachten wir erneut den ernsthaft schlecht gestellten Fall, dass λ beschränkten Träger hat, und erhalten erstaunlicher Weise die exakten Minimax-Raten für beliebige Wahl von γ. Außerdem zeigt sich, dass in dem Spezialfall $\lambda = \mathbf{1}_{[-\eta,\eta]}$ für ein $\eta < \pi/2$ (limited angle) ein einfacher Projektionsschätzer nicht nur effizient sondern sogar exakt adaptiv an die betrachteten Sobolev-Klassen ist.

Part II

Publications

3

Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation

Daniel Hohmann and Hajo Holzmann

Abstract: We show how the multivariate two-component mixtures with independent coordinates in each component by Hall and Zhou (2003) can be studied within the framework of conditional mixtures as recently introduced by Henry, Kitamura and Salanié (2010). Here, the conditional distribution of the random variable Y given the vector of regressors Z can be expressed as a two-component mixture, where only the mixture weights depend on the covariates. Under appropriate tail conditions on the characteristic functions and the distribution functions of the mixture components, which allow for flexible location-scale type mixtures, we show identification and provide asymptotically normal estimators. The main application for our results are bivariate two-component mixtures with independent coordinates, the case not previously covered by Hall and Zhou (2003). In a simulation study we investigate the finite-sample performance of the proposed methods. The main new technical ingredient is the estimation of limits of quotients of two characteristic functions in the tails from independent samples, which might be of some independent interest.

Keywords: characteristic function, conditional mixture, finite mixture, nonparametric estimation

AMS 2000 subject classification: 62G05, 62G20

3.1. Introduction

Finite mixtures are frequently used to model populations with unobserved heterogeneity. While the component distributions are most often chosen from some parametric family, e.g. the normal or t-distributions, cf. McLachlan and Peel (2000), in recent years there has been quite some interest in finite mixtures with nonparametric components, see below for a review of some of the literature.

A prominent example is the multivariate two-component mixture with independent coordinates in each component by Hall and Zhou (2003, HZ in what follows), which they introduced for modeling results of repeated tests on a single person with unknown disease status.

In this paper we show how the model by HZ can be cast into the framework of two-component conditional mixtures by Henry, Kitamura and Salanié (2010, HKS in the following). In particular, our results imply that for the HZ-model in the (in general only partially identified) two-dimensional case, an appropriate representation of the factors in the independent components can still be identified and estimated under some additional tail assumptions.

Suppose that the conditional distribution of the random variable Y given the vector of regressors Z can be expressed as the two-component mixture

$$F(y|z) = (1-\lambda(z))F_0(y) + \lambda(z)F_1(y), \tag{3.1}$$

where only the mixture weights depend on the covariates. Apart from actual dependence of $\lambda(z)$ on the covariates, identification in model (3.1) requires additional assumptions on the component distributions F_0 and F_1. HKS investigate identifiability and estimation under tail conditions of the distribution functions themselves, which are tailored to location-type mixtures, but do not work for scale mixtures. We focus on identification and estimation results of (3.1) under appropriate tail conditions on the characteristic functions of F_0 and F_1, which shall allow for more flexible location-scale-type mixtures. Indeed, our main technical contribution is the derivation of the large-sample theory for characteristic function-based estimators.

Let us review some of the literature on mixtures with nonparametric components. In the simple case of finite mixtures of univariate distributions, most theoretical work assumes symmetry of each component distribution. For example, Bordes, Mottelet and Vandekerkhove (2006) and Hunter, Wang and Hettmansperger (2007) present results on identifiability and asymptotically normal estimation in a two-component location mixture of a single symmetric distribution, and Bordes and Vandekerkhove (2010) and Hohmann and Holzmann (2013a) have similar results in a two-component mixture model with two symmetric components one of which is completely specified while the other is unknown with unknown location parameter. For mixtures of regressions, there is a series of work which exploits the additional information provided by covariates. Kasahara and Shimotsu (2009) extend the HZ-approach to the context of switching regressions. Kitamura (2004, unpublished) considers identifiability issues for univariate mixtures of regressions on the mean functions and obtains full nonparametric identification of the components under some tail assumptions of either their characteristic functions or their moment generating functions. Vandekerkhove (2012) considers the more specific case of a linear switching regression model, where the switching error distributions are only assumed to be symmetric.

Our paper is organized as follows. In Section 3.2 we show how the HZ-model can be studied within the framework of model (3.1). Further, we show identification under tail conditions on the characteristic functions of F_0 and F_1, and in particular conclude that for the HZ-model in the two-dimensional case, an appropriate representation of the component distributions is identified and can be estimated under our assumptions. In Section 3.3 we construct asymptotically normal estimators of the component distributions and the mixture weights. The main new technical ingredient is the estimation of limits of quotients of two characteristic functions in the tails from independent samples, which we discuss in Section 3.4, and which might be of some independent interest. The proofs use strong approximation as well as an entropy-type bound for the characteristic process.

A simulation study is conducted in Section 3.5, where we focus on the HZ-model in two dimensions. We also propose a cross-validation scheme in order to select the tuning parameters of the estimators. Proofs are deferred to an appendix, while some additional technical results are given in the supplementary material Hohmann and Holzmann (2013c) (see Chapter 5).

3.2. Two-component conditional mixtures

In this section, we discuss the model by HZ as our main example of (3.1). Further, we briefly review the identifiability statements of HKS and extend these to cover tail conditions on the characteristic functions of the component distributions.

3.2.1. Examples and model reduction

Example 3.1 (Unobserved binary status)**.** Let Y be some endogenous variable affected by an unobservable binary status $T \in \{0,1\}$. Instead of T we observe the regressor T^* which effects the status variable T, but such that Y given T is independent of T^*. Then

$$\begin{aligned} \mathrm{P}(Y \le y | T^* = z) &= \sum_{s=0,1} \mathrm{P}(T = s | T^* = z)\,\mathrm{P}(Y \le y | T = s) \\ &= (1-\lambda(z))\,F_0(y) + \lambda(z)\,F_1(y) \end{aligned}$$

with $F_j(y) = \mathrm{P}(Y \le y | T = j)$ and $\lambda(z) = \mathrm{P}(T = 1 | T^* = z)$. The main example in HKS is the *misclassified* binary status variable, e.g. the case of binary T^*. □

Example 3.2 (Mixtures with independent components)**.** Suppose that $X_1, \dots, X_k$ are given results of a medical test of a single person where it is unknown if this person is indeed affected by the disease or not. Let T denote the indicator for this unknown affection status, i. e., $T = 1$ if and only if the person is actually diseased. It is reasonable to assume that the k test results are stochastically independent given the status T. Therefore, the observed data can be modeled by the k-variate two-component mixture

$$F(x) = (1-\alpha)\prod_{i=1}^{k} F_{0,i}(x_i) + \alpha \prod_{i=1}^{k} F_{1,i}(x_i), \quad x = (x_1, \dots, x_k)' \in \mathbb{R}^k,$$

with $\alpha = \mathrm{P}(T = 1)$ and $F_{j,i}(x) = \mathrm{P}(X_i \le x | T = j)$. This model was investigated by HZ who established nonparametric identifiability of the cdfs $F_{j,i}$ and the mixture weight α under some mild *irreducibility* condition on F for the case $k \ge 3$. Partial identifiability results were obtained for $k = 2$. Let $i_0 \in \{1, \dots, k\}$ be an arbitrary fixed index, and define

$$Y = X_{i_0}, \qquad Z = (X_1, \dots, X_{i_0-1}, X_{i_0+1}, \dots, X_k)'.$$

Due to the conditional independence of the X_j given T, we obtain for $y \in \mathbb{R}$ and $z \in \mathbb{R}^{k-1}$ that

$$\mathrm{P}(Y \le y | Z = z) = (1-\lambda(z))F_{0,i_0}(y) + \lambda(z)F_{1,i_0}(y), \qquad \lambda(z) = \mathrm{P}(T = 1 | Z = z).$$

Note that the function $\lambda(z)$ is only of minor interest. We may as well condition on $\{Z \in B\}$ for a Borel set $B \subset \mathbb{R}^{k-1}$ to obtain

$$\begin{aligned} \mathrm{P}(Y \le y | Z \in B) &= (1-\pi(B))F_{0,i_0}(y) + \pi(B)F_{1,i_0}(y), \\ \pi(B) &= \mathrm{P}(T = 1 | Z \in B) = \alpha\,\mathrm{P}(Z \in B | T = 1)/\mathrm{P}(Z \in B). \end{aligned} \tag{3.2}$$

In particular, the weight $\pi(\mathbb{R}^{k-1})$ determines the parameter α. □

Motivated by the HZ-model as well as the case of a misclassified binary status variable, we consider conditioning in model (3.1) on events $\{Z \in B\}$ for a Borel set $B \in \mathscr{B}^p$ which have positive probability. Specifically, setting $\pi(B) = E(\lambda(Z)|Z \in B)$ allows to write (3.1) as

$$F(y|B) = (1-\pi(B))F_0(y) + \pi(B)F_1(y). \tag{3.3}$$

3.2.2. Identification

Next, we briefly revisit the results on identification in HKS, reformulated in the context of (3.3), and add a condition for identifiability on the quotients of characteristic functions which allows to identify scale mixtures. As HKS, we start with two basic assumptions.

Assumption A 3.1. 1. There exist $B_0, B_1 \in \mathscr{B}^p$ such that $0 < \pi(B_0), \pi(B_1) < 1$, $\pi(B_0) \neq \pi(B_1)$. 2. There exists a $y_0 \in \mathbb{R}$ such that $F_1(y_0) \neq F_0(y_0)$.

Given Assumption A 3.1 set

$$\xi = \xi(B_0, B_1) = \frac{\pi(B_1)}{\pi(B_0)} \quad \text{and} \quad \zeta = \zeta(B_0, B_1) = \frac{1-\pi(B_1)}{1-\pi(B_0)}.$$

Then direct computations show that

$$\begin{aligned} F_0(y) &= F(y|B_0) + \frac{F(y|B_1) - F(y|B_0)}{1-\xi}, \\ F_1(y) &= F(y|B_0) + \frac{F(y|B_1) - F(y|B_0)}{1-\zeta}. \end{aligned} \tag{3.4}$$

Further, set

$$\Lambda(B) = \frac{F(y_0|B) - F(y_0|B_0)}{F(y_0|B_1) - F(y_0|B_0)} = \frac{\pi(B) - \pi(B_0)}{\pi(B_1) - \pi(B_0)}.$$

Straightforward calculations give $\pi(B_0) = (1-\zeta)/(\xi-\zeta)$ and $\pi(B_1) - \pi(B_0) = -(1-\xi)(1-\zeta)/(\xi-\zeta)$, and thus

$$\pi(B) = \frac{1-\zeta}{\xi-\zeta} - \frac{(1-\xi)(1-\zeta)}{\xi-\zeta}\Lambda(B). \tag{3.5}$$

From (3.4) and (3.5), HKS observe that F_0, F_1 and π (and in particular λ) can be identified from the quantities ξ and ζ and the observable cdf F. So, under Assumption A 3.1, the identification and estimation of ξ and ζ is the crucial part for the mixture (3.3).

In order to achieve full identification, consider the following tail dominance conditions concerning the component cdfs F_0 and F_1 and their Fourier transforms, say $\tilde{F}_0$ and $\tilde{F}_1$, respectively.

C1. $\lim_{y\to-\infty} F_1(y)/F_0(y) = 0$

C2. $\lim_{y\to+\infty} (1-F_0(y))/(1-F_1(y)) = 0$

C3. $\lim_{y\to+\infty} \tilde{F}_0(y)/\tilde{F}_1(y) = 0$

For $i = 2,3$, let $\mathscr{M}_i$ denote the class of mixtures of the form (3.3), the component cdfs of which satisfy the tail conditions C1 and Ci. HKS state identification under C1 and C2. For convenience, we reformulate their result.

Theorem 3.1. *If $F \in \mathscr{M}_i$ for $i = 2$ or $i = 3$, then, under A 3.1, F_0, F_1 and π are nonparametrically identifiable within this class $\mathscr{M}_i$. Moreover, by C1,*

$$\lim_{y \to -\infty} \frac{F(y|B_1)}{F(y|B_0)} = \zeta \tag{3.6}$$

and

$$\text{under C2,} \ \lim_{y \to \infty} \frac{1 - F(y|B_1)}{1 - F(y|B_0)} = \xi, \qquad \text{under C3,} \ \lim_{y \to \infty} \frac{\tilde{F}(y|B_1)}{\tilde{F}(y|B_0)} = \xi. \tag{3.7}$$

Let us comment on and illustrate the conditions C1-C3 as well as the statement of the theorem.

Assumptions C1 and C2 from HKS mean that F_0 dominates the left tail of the distribution F, while F_1 dominates the right tail. This assumption is natural for location mixtures, where it is satisfied for (exponentially) light tails of the underlying distribution. A class of examples is the (skew) normal distribution with equal skewness and scale parameters.

For scale mixtures, Assumptions C1 and C2 are not appropriate. For example, for normal distributions, the component with higher variance dominates both tails. Specifically, consider a normal mixture with $\sigma_0 > \sigma_1$ and $\mu_0, \mu_1 \in \mathbb{R}$ arbitrary. Then Assumptions C1 is satisfied and further

$$\frac{\tilde{F}_0(y)}{\tilde{F}_1(y)} = \exp\big(i(\mu_0 - \mu_1)y\big) \exp\big(-(\sigma_0^2 - \sigma_1^2)x^2/2\big) \longrightarrow 0, \qquad |x| \to \infty,$$

thus Assumptions C1 and C3 provide identification. More generally, for a scale mixture of a supersmooth density (for which the characteristic function decays at an exponential rate), C3 is satisfied. Thus, C3 allows to smoothly separate scale-mixtures.

Example 3.2 *(continued)***.** We show in the context of Example 3.2 that Theorem 3.1 only states that the specific representation of the conditional mixtures for which the components satisfy the conditions C1 and C2 or C3 is identified, there might be further representations of the from (3.3). Nevertheless, as argued above, these representations are quite natural: C1 and C2 are appropriate if F_0 and F_1 dominate distinct tails of the distribution, while C1 and C3 are natural for a scale-type mixture in a smooth density with light tails.

In terms of densities, the model is

$$f(x) = (1-\pi) f_{0,1}(x_1) f_{0,2}(x_2) + \pi f_{1,1}(x_1) f_{1,2}(x_2), \quad x = (x_1, x_2)' \in \mathbb{R}^2. \tag{3.8}$$

Let f_1 and f_2 denote the one-dimensional marginals of f. Then theorem 4.1 in HZ provides the factorization

$$f(x) - f_1(x_1) f_2(x_2) = g_1(x_1) g_2(x_2),$$

where the functions g_1 and g_2 are uniquely determined up to constant multiples. Now, theorem 4.2 in HZ states the partial identifiability of (3.8) as follows. If $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2)$ is a vector of real constants unequal to zero such that $\operatorname{sgn} \alpha_j = -\operatorname{sgn} \beta_j$, the fraction $|\beta_j|/(|\alpha_j| + |\beta_j|)$ does not depend on j, and such that

$$f_{0,j}^{\theta} := f_j + \alpha_j g_j, \quad f_{1,j}^{\theta} := f_j + \beta_j g_j \tag{3.9}$$

are non-negative, then these latter functions are probability densities which also fulfill (3.8), with mixture weight $1 - \pi^{\theta} := |\beta_1|/(|\alpha_1| + |\beta_1|)$. Denote by Θ the set of corresponding shifting vectors θ. We show that our tail assumptions identify a unique value in Θ. Indeed, suppose that $\theta^* \in \Theta$ is such that

$$0 = \lim_{x\to-\infty} \frac{f_{1,1}^{\theta^*}(x)}{f_{0,1}^{\theta^*}(x)} = \lim_{x\to-\infty} \frac{1+\beta_1^* g_1(x)/f_1(x)}{1+\alpha_1^* g_1(x)/f_1(x)}, \tag{3.10}$$

corresponding to condition C1 of Section 3.2 in terms of densities. Since $\beta_1^* \neq 0$, this implies the convergence

$$\lim_{x\to-\infty} g_1(x)/f_1(x) = -1/\beta_1^*,$$

which by (3.8) yields the identification of f_1. The arguments under C2 and C3 are similar. □

The nonidentifiability without additional assumptions on the components in model (3.3) holds more generally, see Example 5.1 in Hohmann and Holzmann (2013c). Further, without the additional regressor Z even in case of a known weight and tail conditions on the components, the model is not identified, see Example 5.2 in Hohmann and Holzmann (2013c).

3.3. Estimation

In this section, given i.i.d. observations $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ such that the conditional distribution $F(y|z)$ satisfies (3.3), following HKS we propose nonparametric estimators of the components F_0, F_1 and of the weight function π. The essential step is to estimate the quantities ζ and ξ, then, based on (3.4) and (3.5), plug-in estimates are easily devised. Given the tail conditions C1 and C2/C3, we estimate ζ and ξ as the limits arising in (3.6) and (3.7). To this end, Section 3.4 contains the asymptotic distribution theory for limits of quotients of characteristic functions, our major contribution, as well as of distribution functions, which is essentially covered in HKS, of independent samples in their tails. Here, we apply this theory to obtain asymptotics for the estimators in model (3.3).

3.3.1. Estimation of ζ and ξ under C1 and C3

Consider the empirical conditional distribution and characteristic functions

$$F_n(y|B_j) = \frac{\sum_{k=1}^n \mathbf{1}_{\{Y_k \le y\}} \mathbf{1}_{\{Z_k \in B_j\}}}{\sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_j\}}}, \qquad y \in \mathbb{R},$$

$$\widetilde{F}_n(t|B_j) = \frac{\sum_{k=1}^n \exp(itY_k) \mathbf{1}_{\{Z_k \in B_j\}}}{\sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_j\}}}, \qquad t \in \mathbb{R}.$$

Motivated by (3.6) and (3.7), following HKS we consider estimators for ζ and ξ of the form

$$\zeta_n = \frac{F_n(L_n|B_1)}{F_n(L_n|B_0)}, \qquad \xi_n = \operatorname{Re} \frac{\widetilde{F}_n(R_n|B_1)}{\widetilde{F}_n(R_n|B_0)},$$

where the levels L_n and R_n need to be chosen appropriately. To this end, assume

Assumption A 3.2. The Borel sets B_0 and B_1 in A 3.1 i. satisfy $B_0 \cap B_1 = \emptyset$ and $p_j := \mathrm{P}(Z \in B_j) > 0$, $j = 0, 1$.

Under A 3.2, we define disjoint subsamples $Y_1^*, \dots, Y_{m_n}^*$ and $Y_1^{**}, \dots, Y_{l_n}^{**}$ of $Y_1, \dots, Y_n$, where the Y_j^* correspond to the observations Y_k such that $Z_k \in B_0$, the Y_j^{**} correspond to $Z_k \in B_1$, and set $m_n = \sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_0\}}$ and $l_n = \sum_{k=1}^n \mathbf{1}_{\{Z_k \in B_1\}}$.

We choose L_n as an intermediate lower order statistic of the subsample $Y_1^*, \dots, Y_{m_n}^*$: Let L_n be the $\lfloor r_n \rfloor$-th largest order statistic of the Y_k^*, $L_n = Y^*_{m_n(\lfloor r_n \rfloor)}$, where $r_n \to \infty$ is such that

$$r_n/n \to 0, \quad r_n/\sqrt{n} \to \infty \qquad \text{as } n \to \infty. \tag{3.11}$$

In order to choose R_n, we assume that the characteristic function satisfies $\widetilde{F}(t|B_0) \to 0$ as $t \to \infty$. Let $s_n \to \infty$ be such that $s_n/n \to 0$. Then, for large n, by continuity of $\widetilde{F}(y|B_0)$ there is a (not necessarily unique) solution t_n of the equation $\widetilde{F}(t_n|B_0) = \sqrt{s_n/m_n}$, and we choose R_n as a solution of the corresponding empirical version of this equation: $|\widetilde{F}_n(R_n|B_0)| = \sqrt{s_n/m_n}$. More precisely, we require

Assumption A 3.3. There exists $\gamma > 0$ and a non-random sequence $t_n \to \infty$ such that (3.20)-(3.23) (see Section 3.4) hold true for $F = F(\cdot|B_1)$, $G = F(\cdot|B_0)$, and $h_n = R_n$.

For further discussion see Section 3.4. Finally, assume that the rates r_n and s_n are chosen such that there exist constants β_ζ and β_ξ in $\mathbb{R}$ (possibly zero) for which

$$\sqrt{r_n}\Big(\frac{F(L_n|B_1)}{F(L_n|B_0)} - \zeta\Big) \to_p \beta_\zeta, \qquad \sqrt{s_n}\Big(\operatorname{Re}\frac{\widetilde{F}(R_n|B_1)}{\widetilde{F}(R_n|B_0)} - \xi\Big) \to_p \beta_\xi \tag{3.12}$$

as $n \to \infty$.

Proposition 3.1. *If A 3.1 - A 3.3 and (3.12) hold, then*

$$\sqrt{r_n}(\zeta_n - \zeta) \rightsquigarrow \mathcal{N}\big(\beta_\zeta, \tau\zeta + \zeta^2\big), \qquad \sqrt{s_n}(\xi_n - \xi) \rightsquigarrow \mathcal{N}\big(\beta_\xi, (\tau + \xi^2)/2\big)$$

with $\tau = p_0/p_1$. Moreover, if $s_n = r_n$, then the estimators are asymptotically independent.

The first part of the proposition is as in HKS, the second (which is based on characteristic functions) as well as the asymptotic independence are our main contributions to estimation. Under the Assumptions C1 and C2, we obtain analogous results to those in HKS, see Hohmann and Holzmann (2013c) for the details.

For further discussion of the tail assumption (3.12) see HKS (Lemmas 7 and 8). Note that it also applies to the characteristic functions if these (and not the distributions functions) satisfy the shape constraints made in HKS.

3.3.2. Estimating the component distributions and the weight function

We now turn to the estimation of the components distributions F_0 and F_1 and the mixture weight function π. We obtain similar, though slightly more refined results as HKS.

By (3.4), natural estimates of F_0 and F_1 are given by

$$\hat{F}_0(y) = F_n(y|B_0) + \frac{F_n(y|B_1) - F_n(y|B_0)}{1-\xi_n},$$
$$\hat{F}_1(y) = F_n(y|B_0) + \frac{F_n(y|B_1) - F_n(y|B_0)}{1-\zeta_n},$$

where ζ_n is obtained using C1 and ξ_n either from C2 or from C3. As a consequence of Proposition 3.1,

$$\sqrt{r_n}(\zeta_n - \zeta) \rightsquigarrow \mathcal{N}\left(\beta_\zeta, \sigma_\zeta^2\right), \qquad \sqrt{s_n}(\xi_n - \xi) \rightsquigarrow \mathcal{N}\left(\beta_\xi, \sigma_\xi^2\right), \tag{3.13}$$

where the variances σ_ζ^2 and σ_ξ^2 are given as in Proposition 3.1, and possibly $r_n = s_n$, in which case the estimators are asymptotically independent. Then we have

Theorem 3.2. *Suppose that (3.13) holds, where s_n and r_n satisfy (3.11). Then*

$$\sqrt{s_n}\big(\hat{F}_0(y) - F_0(y)\big) \overset{\ell^\infty(\mathbb{R})}{\rightsquigarrow} \mathbf{G}_\xi(y), \qquad \sqrt{r_n}\big(\hat{F}_1(y) - F_1(y)\big) \overset{\ell^\infty(\mathbb{R})}{\rightsquigarrow} \mathbf{G}_\zeta(y),$$

where $\mathbf{G}_i$, $i = \xi, \zeta$, are tight Gaussian processes with mean and covariance functions

$$\mu_i(y) = D(F,y)(1-i)^{-2}\beta_i, \quad \rho_i(y_1,y_2) = D(F,y_1)D(F,y_2)(1-i)^{-4}\sigma_i^2,$$

$i = \xi, \zeta$, where $D(F,y) = F(y|B_1) - F(y|B_0)$.

We note that relations analogous to (3.4) also hold true for underlying densities, and hence that similar estimators for the densities could be devised.

Finally, we consider estimation of the mixture weight function $\pi(B)$ for sets $B \in \mathscr{B}^p$ with $\mathrm{P}(Z \in B) > 0$, $B = \mathbb{R}^p$ being of particular interest. Fix a y_0 satisfying Assumption A 3.1. From (3.5), a suitable estimator is given by

$$\pi_n(B) = L_1(\xi_n,\zeta_n) - L_2(\xi_n,\zeta_n)\frac{F_n(y_0|B) - F_n(y_0|B_0)}{F_n(y_0|B_1) - F_n(y_0|B_0)},$$

where $L_1(x_1,x_2) = (1-x_2)/(x_1-x_2)$, $L_2(x_1,x_2) = (1-x_1)(1-x_2)/(x_1-x_2)$, $x_1 \neq x_2$.

Theorem 3.3. *Let $B \in \mathscr{B}^p$ with $\mathrm{P}(Z \in B) > 0$, and assume (3.13), where in case of equal rates we additionally assume asymptotic independence.*

1. If $s_n = r_n$, we have

$$\sqrt{r_n}\big(\pi_n(B) - \pi(B)\big) \rightsquigarrow \mathcal{N}\big((\beta_\xi,\beta_\zeta)\mathrm{J}, \mathrm{J}'\operatorname{diag}(\sigma_\xi^2,\sigma_\zeta^2)\mathrm{J}\big),$$

where

$$\mathrm{J} = (\xi-\zeta)^{-2}\left(\begin{pmatrix}\zeta-1\\1-\xi\end{pmatrix} - \Lambda(B)\begin{pmatrix}-(1-\zeta)^2\\(1-\xi)^2\end{pmatrix}\right).$$

2. If $s_n = o(r_n)$, then

$$\sqrt{s_n}\big(\pi_n(B) - \pi(B)\big) \rightsquigarrow \mathcal{N}\big(\beta_\xi \mathrm{j}, \mathrm{j}^2\sigma_\xi^2\big), \quad \mathrm{j} = (\xi-\zeta)^{-2}\big(\zeta - 1 + \Lambda(B)(1-\zeta)^2\big).$$

Remark. When estimating the mixture weights $\pi(B_0)$ and $\pi(B_1)$ of the sets B_0 and B_1 upon which the estimation procedure is based, the asymptotic covariance matrix has a simpler form: Since $\Lambda(B_0) = 0$ and $\Lambda(B_1) = 1$, we obtain that

$$\mathrm{J}(B_0, B_1) = \frac{1}{(\xi - \zeta)^2} \begin{pmatrix} -(1-\zeta) & 1-\xi \\ -\zeta(1-\zeta) & \xi(1-\xi) \end{pmatrix}.$$

3.4. Estimating quotients in the tails

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be mutually independent sequences of i.i.d. observations with distribution functions F and G, respectively. Assume that

$$F(y)/G(y) \longrightarrow \theta \qquad \text{as } y \to -\infty \tag{3.14}$$

or / and

$$\widetilde{F}(y), \widetilde{G}(y) \longrightarrow 0, \quad \widetilde{F}(y)/\widetilde{G}(y) \longrightarrow \eta \qquad \text{as } y \to \infty \tag{3.15}$$

hold for some $\theta > 0$ and $\eta \in \mathbb{C} \backslash \{0\}$, where as above $\widetilde{F}$ and $\widetilde{G}$ denote the characteristic functions of F and G. We shall construct asymptotically normal estimators of θ and η. In the following, suppose that l_n and m_n are sequences in $\mathbb{N}$ such that $l_n, m_n \asymp n$ as $n \to \infty$.

3.4.1. Characteristic Functions

To estimate η in (3.15), let

$$\eta_n = \widetilde{F}_n(h_n)/\widetilde{G}_n(h_n),$$

where

$$\widetilde{F}_n(y) = \frac{1}{l_n} \sum_{k=1}^{l_n} \exp(iyX_k), \qquad \widetilde{G}_n(y) = \frac{1}{m_n} \sum_{k=1}^{m_n} \exp(iyY_k),$$

with h_n a sequence tending to infinity. Decompose

$$\eta_n - \eta = (\eta_n - \bar{\eta}_n) + (\bar{\eta}_n - \eta), \qquad \bar{\eta}_n = \widetilde{F}(h_n)/\widetilde{G}(h_n).$$

In order to handle the "variance term", write

$$\sqrt{s_n}(\eta_n - \bar{\eta}_n) = \frac{\sqrt{s_n/m_n}}{\widetilde{G}_n(h_n)} \Big(\sqrt{m_n/l_n}\, \widetilde{\mathbf{F}}_n(h_n) - \bar{\eta}_n \widetilde{\mathbf{G}}_n(h_n) \Big), \tag{3.16}$$

where $\widetilde{\mathbf{F}}_n = \sqrt{l_n}(\widetilde{F}_n - \widetilde{F})$ and $\widetilde{\mathbf{G}}_n = \sqrt{m_n}(\widetilde{G}_n - \widetilde{G})$ are the characteristic processes and $s_n \to \infty$. Assume that s_n satisfies

$$s_n/n \to 0 \qquad \text{as } n \to \infty, \tag{3.17}$$

and that $h_n \to_p \infty$ is chosen such that

$$|\widetilde{G}_n(h_n)| = \sqrt{s_n/m_n}(1 + o_P(1)). \tag{3.18}$$

We shall use strong approximations of the characteristic processes by

$$\mathbf{C}(y) = \int \exp(iyx)\,\mathbf{B}(F(dx)) \tag{3.19}$$

for $\widetilde{\mathbf{F}}_n(y)$, and similarly for $\widetilde{\mathbf{G}}_n$. In order that these processes are sample-continuous and that strong approximations work, some conditions on F and G are required, see Csörgő (1981). We shall adopt the following sufficient condition: Assume that there exists $\gamma > 0$ such that

$$y^\gamma H(-y) + y^\gamma(1-H(y)) = O(1) \quad \text{as } y \to \infty, \quad H = F, G. \tag{3.20}$$

Finally, we assume that there also exists a non-random sequences $t_n \to \infty$ such that

$$t_n = o\big(n^{\gamma/(2\gamma+4)}(\log n)^{-(\gamma+1)/(\gamma+2)}\big), \tag{3.21}$$

$$|h_n - t_n| = o_P(1), \tag{3.22}$$

$$|\widetilde{G}(h_n) - \widetilde{G}(t_n)| = o_P\big(\sqrt{s_n/m_n}\big), \tag{3.23}$$

with γ determined by (3.20).

Remark. Given a non-random sequence $t_n \to \infty$ of order (3.21), assume that the following separability criterion holds: There is a sequence a_n, either constant or tending to infinity, such that for all $\varepsilon > 0$ there exists $C_\varepsilon > 0$ fulfilling

$$\inf_{a_n|t-t_n|>\varepsilon} \big|\sqrt{m_n/s_n}\,|\widetilde{G}(t)| - 1\big| > C_\varepsilon \tag{3.24}$$

for n sufficiently large. In case of a supersmooth density, one can show that (3.24) holds with $a_n = t_n$, and that this rate implies (3.23) if t_n is chosen such that $|\widetilde{G}(t_n)| = \sqrt{s_n/m_n}$.

Lemma 3.1. *Assume that (3.20) and (3.24) hold. If t_n is a non-random sequence of order (3.21), then there exists a sequence h_n such that $\widetilde{G}_n(h_n) = \sqrt{s_n/m_n}$ and $|h_n - t_n| = o_P(a_n^{-1})$.*

Example 3.3. Consider exemplarily the Gaussian characteristic function

$$\widetilde{\Phi}(x) = \widetilde{\Phi}(x;\mu,\sigma^2) = \exp(i\mu x)\exp(-\sigma^2 x^2/2),$$

and let t_n be chosen such that $|\widetilde{\Phi}(t_n)| = \sqrt{s_n/m_n}$. Given $\varepsilon > 0$, the infimum in (3.24) is attained at $t_n^* = t_n + \varepsilon/a_n$. A Taylor expansion then yields

$$\big|\sqrt{m_n/s_n}|\widetilde{\Phi}(t_n^*) - 1\big| = \sqrt{m_n/s_n}\,\sigma^2\bar{t}_n \exp(-\sigma^2\bar{t}_n^2/2)\,\varepsilon/a_n$$

with $\bar{t}_n \in [t_n, t_n + \varepsilon/a_n]$. Choosing $a_n = t_n$, it follows that $\bar{t}_n/a_n \to 1$ and

$$\exp(-\sigma^2\bar{t}_n^2/2) \geq \exp(-\sigma^2(t_n+\varepsilon/a_n)^2/2) = \sqrt{s_n/m_n}\exp\big(-\sigma^2(2\varepsilon + (\varepsilon/a_n)^2)/2\big),$$

and thus (3.24) holds. As a result, by Lemma 3.1 there exists a random sequence h_n such that $|\widetilde{\Phi}_n(h_n)| = \sqrt{s_n/m_n}$ and $|h_n - t_n| = o_P(t_n^{-1})$. Now, again by a Taylor expansion, similar arguments show that for some $\bar{t}_n$ between h_n and t_n,

$$\begin{aligned}\sqrt{m_n/s_n}\big|\operatorname{Re}\Phi(h_n) - \operatorname{Re}\Phi(t_n)\big| &= \sqrt{m_n/s_n}\big(\sigma^2\bar{t}_n\cos(\mu\bar{t}_n) + \mu\sin(\mu\bar{t}_n)\big) \\ &\quad \exp(-\sigma^2\bar{t}_n^2/2)|h_n - t_n| \\ &= o_P(1).\end{aligned}$$

Since the imaginary part of $\widetilde{\Phi}(h_n) - \widetilde{\Phi}(t_n)$ can be handled likewise, we also see that (3.23) is fulfilled. □

Theorem 3.4. *Assume that (3.17), (3.18) and (3.20)-(3.23) hold. If there exists $\tau > 0$ such that $m_n/l_n \to \tau$, then*

$$\sqrt{s_n}(\eta_n - \bar{\eta}_n) \rightsquigarrow \mathcal{CN}(0, \tau + |\eta|^2, 0),$$

where $\mathcal{CN}$ denotes the complex-normal distribution. More explicitly,

$$\sqrt{s_n}\begin{pmatrix}\mathrm{Re}(\eta_n - \bar{\eta}_n)\\ \mathrm{Im}(\eta_n - \bar{\eta}_n)\end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \frac{1}{2}\begin{pmatrix}\tau + |\eta|^2 & 0\\ 0 & \tau + |\eta|^2\end{pmatrix}\right).$$

3.4.2. Distribution functions

To estimate θ in (3.14), let

$$\theta_n = F_n(h_n)/G_n(h_n), \quad F_n(y) = \frac{1}{l_n}\sum_{k=1}^{l_n} \mathbf{1}_{\{X_k \le y\}}, \quad G_n(y) = \frac{1}{m_n}\sum_{k=1}^{m_n} \mathbf{1}_{\{Y_k \le y\}},$$

where the level h_n is specified below. Write

$$\theta_n - \theta = (\theta_n - \bar{\theta}_n) + (\bar{\theta}_n - \theta), \qquad \bar{\theta}_n = F(h_n)/G(h_n).$$

Assume that $r_n \to \infty$ satisfies (3.11), and that $h_n \to_p -\infty$ is chosen such that

$$G_n(h_n) = r_n/m_n + o_P(r_n/n) = r_n/m_n\,(1 + o_P(1)). \tag{3.25}$$

(3.25) is satisfied if we choose in particular $h_n = Y_{m_n(\lfloor r_n \rfloor)}$, where $\lfloor r_n \rfloor$ is the largest integer smaller than r_n, and where $Y_{m_n(\lfloor r_n \rfloor)}$ denotes the $\lfloor r_n \rfloor$-th largest order statistic of the sample $Y_1, \ldots, Y_{m_n}$, since $G_n(h_n) = \lfloor r_n \rfloor / m_n = r_n/m_n(1 + o(1))$.

Theorem 3.5. *Suppose that the assumptions of Theorem 3.4 for $s_n = r_n$ as well as (3.11) and (3.25) hold. If there exists $\tau > 0$ such that $m_n/l_n \to \tau$, then*

$$\sqrt{r_n}\begin{pmatrix}\theta_n - \bar{\theta}_n\\ \mathrm{Re}\left(\eta_n - \bar{\eta}_n\right)\\ \mathrm{Im}\left(\eta_n - \bar{\eta}_n\right)\end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \frac{1}{2}\begin{pmatrix}2(\tau\theta + \theta^2) & 0 & 0\\ 0 & \tau + |\eta|^2 & 0\\ 0 & 0 & \tau + |\eta|^2\end{pmatrix}\right).$$

The asymptotic distribution of $\sqrt{r_n}(\theta_n - \bar{\theta}_n)$ follows from arguments along the lines of HKS, however, the asymptotic independence requires some additional work. Further applications, discussed in Hohmann and Holzmann (2013c), are testing for tail dominance, as well as estimating the exponent of regular variation.

3.5. Simulation study

We investigate the finite-sample performance of the estimators in a simulation study in the HZ-model. Consider a random vector $(Y, Z)'$ distributed according to

$$F_{Y,Z}(y,z) = (1-p)\,\Psi(y; \mu_0, \sigma_0^2, \lambda_0)\,\Phi(z; 0, 1) + p\,\Psi(y; \mu_1, \sigma_1^2, \lambda_1)\,\Phi(z; 3, 1),$$

where Φ and Ψ denote the normal and skew-normal distribution functions, resp. The distribution of Y conditional on Z is then given by (cf. (3.2))

$$F(y|B) = (1-\pi(B))\,\Psi(y;\mu_0,\sigma_0^2,\lambda_0) + \pi(B)\,\Psi(y;\mu_1,\sigma_1^2,\lambda_1). \tag{3.26}$$

As true values we choose $(\mu_0,\sigma_0^2,\lambda_0) = (0,1,2)$, $(\mu_1,\sigma_1^2,\lambda_1) = (0.5,0.5,3)$, and $p = 0.6$. Let us show that the component distributions in (3.26) then fulfill the tail dominance assumptions C1 and C3.

The density of the skew-normal distribution is given by

$$\psi(x;\mu,\sigma^2,\lambda) = \frac{2}{\sigma}\,\phi\Big(\frac{x-\mu}{\sigma}\Big)\Phi\Big(\lambda\,\frac{x-\mu}{\sigma}\Big), \qquad x\in\mathbb{R},$$

where ϕ denotes the standard normal density, and its characteristic function by

$$\widetilde{\Psi}(x;\mu,\sigma^2,\lambda) = e^{i\mu x-\frac{1}{2}\sigma^2x^2}\Big(1+i\,\mathfrak{I}\Big(\sigma x\lambda/\sqrt{1+\lambda^2}\Big)\Big),$$

where $\mathfrak{I}(x) = \int_0^x \sqrt{2/\pi}\,e^{u^2/2}\,du$. Hence,

$$\Psi(x;\mu,\sigma^2,\lambda) \sim \frac{1}{\lambda\pi(\mu-x)}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2(1+\lambda^2)}, \qquad x\to-\infty,$$

and, as $x\to\infty$,

$$\begin{aligned}|\widetilde{\Psi}(x;\mu,\sigma^2,\lambda)| &= e^{-\sigma^2x^2/2}\sqrt{1+\mathfrak{I}\Big(\sigma x\lambda/\sqrt{1+\lambda^2}\Big)^2}\\ &\sim \sqrt{\frac{2}{\pi}}\,\frac{\sqrt{1+\lambda^2}}{\lambda\sigma x}\,e^{-\frac{1}{2}\sigma^2x^2(1+\lambda^2)^{-1}},\end{aligned}$$

and thus the condition

$$\sigma_0^2/\sigma_1^2 > (1+\lambda_0^2)/(1+\lambda_1^2)$$

is sufficient and necessary for both C1 and C3 to hold true.

For the estimation we further chose $B_0 = (0,\infty)$ and $B_1 = (-\infty,0]$, inducing true values $\zeta = 2.3876$ and $\xi = 0.4502$, and set $y_0 = 1$. The estimation results for ξ, ζ, and p, using different sample sizes n and different rates $r_n = s_n = n^\delta$, are presented in Table 3.1 (a)-(c).

The choice of r_n turns out to highly affect the estimates' variance and bias properties. A small r_n leads to small bias, it however increases the variance, as should be expected from the theory. Therefore, in a second step we use a cross-validation scheme to choose r_n. This can for example be done by a repeated random sub-sampling validation, i. e., one randomly slits up the sample of observations into two sub-samples of equal size, uses these sub-samples to estimate both the mixture

$$(1-\hat{p})\,\hat{F}_0(x) + \hat{p}\,\hat{F}_1(x),$$

for the given r_n, where $\hat{p} = \pi_n(\mathbb{R})$, and the ordinary empirical distribution function of Y. One estimates the L_1-distance between these estimates, the cross-validated r_n is then the minimizer (on some fixed grid). The estimates for ξ, ζ, and p using cross-validation can be found in Table 3.1 (d). Also, Figure 3.1 shows estimates of the distribution functions F_0 and F_1 using cross-validation.

Table 3.1: Estimates of ζ, ξ, and π in the model of Hall and Zhou (2003) for true values $\zeta = 2.3876$, $\xi = 0.4502$, and $p = 0.6$. The table shows mean values and standard deviations (in brackets) of 10^4 repetitions, based on different sample sizes n and different rates $r_n = s_n = n^{\delta}$.

(a) Estimates using $\delta = 0.4$.

n	ζ_n	ξ_n	p_n
500	2.3346 (0.9088)	0.3634 (0.3974)	0.5266 (0.2544)
1000	2.3736 (0.8536)	0.3626 (0.3433)	0.5308 (0.2311)
10000	2.3826 (0.5371)	0.3995 (0.2089)	0.5649 (0.1321)

(b) Estimates using $\delta = 0.6$.

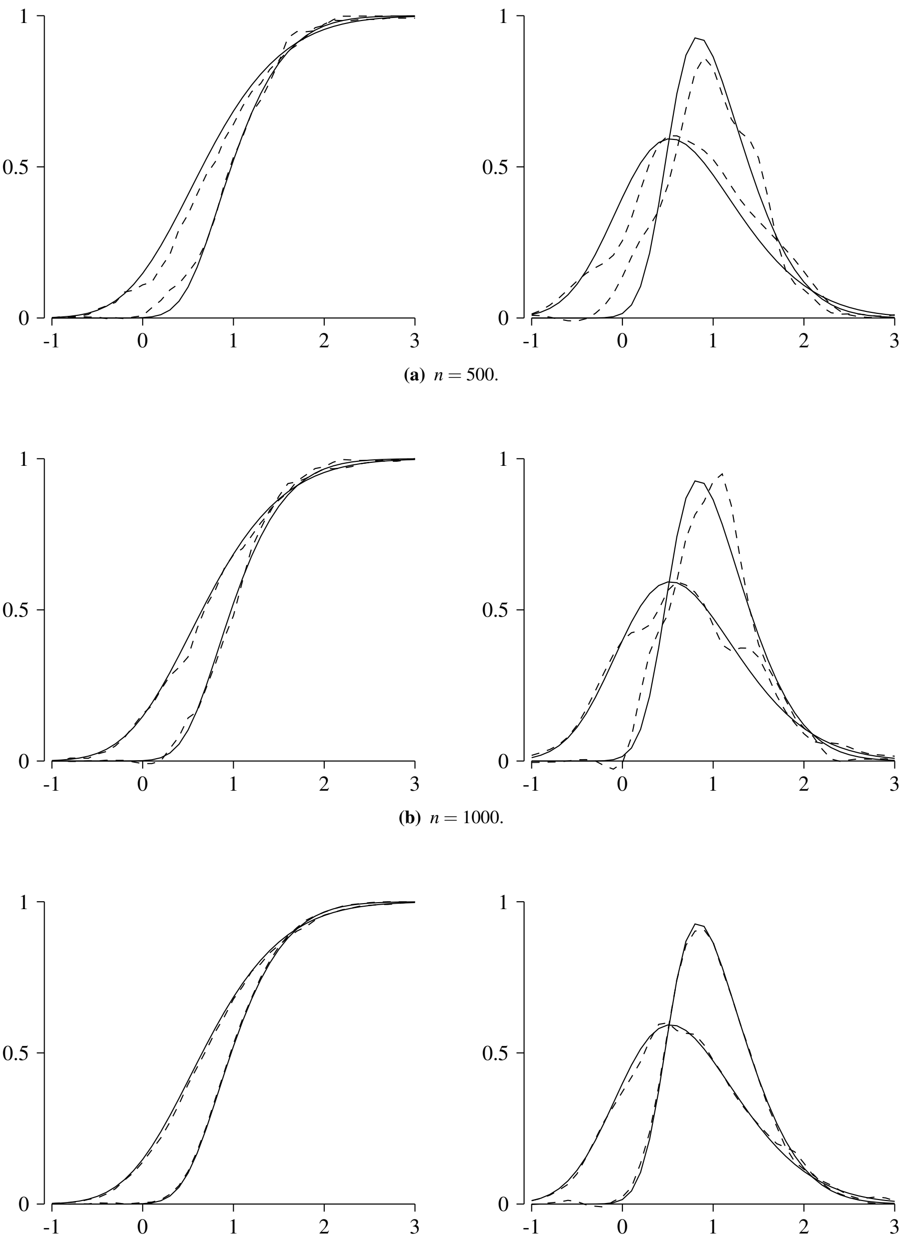
n	ζ_n	ξ_n	p_n
500	1.9299 (0.3493)	0.5129 (0.2249)	0.5685 (0.1319)
1000	2.0911 (0.3201)	0.4755 (0.1825)	0.5734 (0.0976)
10000	2.3657 (0.2009)	0.3967 (0.0853)	0.5693 (0.0495)

(c) Estimates using $\delta = 0.8$.

n	ζ_n	ξ_n	p_n
500	1.2328 (0.1147)	0.7408 (0.1128)	0.4356 (0.1725)
1000	1.2935 (0.0935)	0.7163 (0.0876)	0.4663 (0.1029)
10000	1.5335 (0.0440)	0.6305 (0.0359)	0.5266 (0.0241)

(d) Estimates using cross validation for δ.

n	ζ_n	ξ_n	p_n	δ_n
500	2.2586 (0.8189)	0.3827 (0.3668)	0.5301 (0.2027)	0.4675 (0.0942)
1000	2.2757 (0.6341)	0.3816 (0.2631)	0.5392 (0.1341)	0.5181 (0.0863)
10000	2.3657 (0.2652)	0.3894 (0.0983)	0.5649 (0.0528)	0.5772 (0.0552)



(a) $n = 500$.

(b) $n = 1000$.

(c) $n = 10000$.

Figure 3.1: Estimates $\hat{F}_0$ and $\hat{F}_1$ (dashed) of the component distribution functions F_0 and F_1 (solid) for different sample sizes n. The right column shows estimates of the corresponding densities.

Acknowledgements

The authors gratefully acknowledge financial support from the DFG, grant Ho 3260/3-1. Comments from the associate editor and a referee helped to improve the presentation of the paper.

3.6. Appendix: Proofs

Proofs of Section 3.4

Proof of Lemma 3.1. Let $\delta > 0$, and without loss of generality assume that $x \mapsto |\widetilde{G}(x)|$ is non-increasing for x sufficiently large. Then, by (3.24) there exists $0 < C_\delta < 1$ such that

$$|\widetilde{G}(t_n+\delta)| \le (1-C_\delta)\sqrt{s_n/m_n}\,.$$

From Lemma 3.2 and (3.29) it follows that $|\widetilde{G}_n(t_n+\delta) - \widetilde{G}(t_n+\delta)| = o_P(\sqrt{s_n/m_n})$, yielding

$$\big||\widetilde{G}_n(t_n+\delta)| - |\widetilde{G}(t_n+\delta)|\big| < C_\delta/2 + o_P(1)$$

and thus

$$|\widetilde{G}_n(t_n+\delta)| < \sqrt{s_n/m_n} + o_P(1)\,.$$

As a result, there exists a sequence h_n such that $\widetilde{G}_n(h_n) = \sqrt{s_n/m_n}$, where h_n fulfills the order condition (3.21), too. Hence, it also holds that $|\widetilde{G}_n(h_n) - \widetilde{G}(h_n)| = o_P(\sqrt{s_n/m_n})$, yielding

$$\big|\sqrt{m_n/s_n}|\widetilde{G}(h_n)| - 1\big| = \sqrt{m_n/s_n}\big||\widetilde{G}_n(h_n)| - |\widetilde{G}(h_n)|\big| = o_P(1)\,.$$

Since, for all $\varepsilon > 0$, (3.24) implies

$$\big\{a_n|h_n - t_n| > \varepsilon\big\} \subset \big\{\big|\sqrt{m_n/s_n}|\widetilde{G}(h_n)| - 1\big| > C_\varepsilon\big\}\,,$$

we conclude that $|h_n - t_n| = o_P(a_n^{-1})$. □

Proof of Theorem 3.4. The proof of Theorem 3.4 proceeds in several steps.

Lemma 3.2. *Assume that (3.20) holds. On a sufficiently rich probability space there exist versions of the X_k and Y_k, and independent sequences $\mathbf{B}_{1,n}$ and $\mathbf{B}_{2,n}$ of standard Brownian bridges on $[0,1]$ such that, defining*

$$\mathbf{C}_{1,n}(y) = \int \exp(iyx)\,\mathbf{B}_{1,n}(F(dx)), \quad \mathbf{C}_{2,n}(y) = \int \exp(iyx)\,\mathbf{B}_{2,n}(G(dx)), \tag{3.27}$$

for all sequences $T_n = o\big(n^{\gamma/(2\gamma+4)}(\log n)^{-(\gamma+1)/(\gamma+2)}\big)$ it holds that

$$\sup_{0\le y\le T_n}\big|\widetilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)\big| \to_p 0, \qquad \sup_{0\le y\le T_n}\big|\widetilde{\mathbf{G}}_n(y) - \mathbf{C}_{2,m_n}(y)\big| \to_p 0.$$

Proof. According to Theorem 4 and Corollary 2 in Csörgő (1981), for all $T, \delta > 0$ and $n \in \mathbb{N}$ there exists a constant C which depends only on δ and F such that

$$\mathrm{P}\Big(\sup_{0\le y\le T} \big|\widetilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)\big| > Cq_nT\Big) \lesssim Tn^{-(1+\delta)},$$

where the sequence q_n satisfies $q_n \sim n^{-\gamma/(2\gamma+4)}(\log n)^{(\gamma+1)/(\gamma+2)}$. Hence, for all sequences $T_n = o(q_n^{-1})$ and all $\varepsilon > 0$ we find that

$$\mathrm{P}\Big(\sup_{0\le y\le T_n} \big|\widetilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)\big| > \varepsilon\Big) \lesssim T_n n^{-(1+\delta)}$$

eventually, where the right side converges to zero as $n \to \infty$. □

Remark. The processes $\mathbf{C}_{i,n}$ are zero mean complex Gaussian. In particular, with $\mathbf{C}$ being defined as in (3.19) and $\mathbf{W}$ a standard Brownian motion on $[0,1]$,

$$\mathbf{C}(t) =_d \int_0^1 \exp\big(itF^{-1}(x)\big)\,\mathbf{W}(dx) - \mathbf{W}(1)\int_0^1 \exp\big(itF^{-1}(x)\big)\,dx.$$

With this, basic properties of the Itô integral give

$$\begin{aligned}
\mathrm{E}\big(\mathbf{C}(s)\overline{\mathbf{C}(t)}\big) &= \int \exp(isx)\overline{\exp(itx)}\,F(dx) - \int \exp(isx)\,F(dx)\int \overline{\exp(itx)}\,F(dx) \\
&= \widetilde{F}(s-t) - \widetilde{F}(s)\widetilde{F}(-t), \qquad (3.28)\\
\mathrm{E}\big(\mathbf{C}(s)\mathbf{C}(t)\big) &= \int \exp(isx)\exp(itx)\,F(dx) - \int \exp(isx)\,F(dx)\int \exp(itx)\,F(dx) \\
&= \widetilde{F}(s+t) - \widetilde{F}(s)\widetilde{F}(t),
\end{aligned}$$

and hence $\mathbf{C}(t)$ has variance $\sigma^2(t) = \mathrm{E}\big(|\mathbf{C}(t)|^2\big) = 1 - |\widetilde{F}(t)|^2$ and relation $\rho(t) = \mathrm{E}\big(\mathbf{C}(t)^2\big) = \widetilde{F}(2t) - \widetilde{F}(t)^2$. In particular, if $t_n \to \infty$, then

$$\mathbf{C}(t_n) \rightsquigarrow \mathcal{CN}(0,1,0) \quad \text{as } n \to \infty. \qquad (3.29)$$

Lemma 3.3. *Let $\mathbf{B}_n$ be a sequence of standard Brownian bridges, not necessarily independent of h_n, and define $\mathbf{C}_n(y) = \int \exp(iyx)\,\mathbf{B}_n(F(dx))$. If (3.20) and (3.22) hold, then $|\mathbf{C}_n(h_n) - \mathbf{C}_n(t_n)| = o_P(1)$.*

Proof. For any $\varepsilon, \delta > 0$, defining $I_{n,\delta} = [t_n - \delta, t_n + \delta]$,

$$\mathrm{P}(|\mathbf{C}_n(h_n) - \mathbf{C}_n(t_n)| > \varepsilon) \le \mathrm{P}\Big(\sup_{t\in I_{n,\delta}} |\mathbf{C}_n(t) - \mathbf{C}_n(t_n)| > \varepsilon\Big) + \mathrm{P}(|h_n - t_n| > \delta).$$

The right probability tends to zero due to (3.22). The left probability can be made arbitrarily small by the choice of δ. In fact, by the maximal inequality as given in Lemma 2.1 in Talagrand (1996), there exists a finite constant K such that, for all $x > 0$,

$$\mathrm{P}\Big(\sup_{t\in I_{n,\delta}} |\mathbf{C}_n(t) - \mathbf{C}_n(t_n)| > Kx\int_0^\infty \sqrt{\log N(I_{n,\delta},\eta)}\,d\eta\Big) \le \exp(-x^2),$$

where $N(T,\eta)$ denotes the smallest number of open balls of radius η with respect to the distance

$d(s,t) = [\mathrm{E}|\mathbf{C}_n(s) - \mathbf{C}_n(t)|^2]^{1/2}$ (which does not depend on n) that are necessary in order to cover an index set $T \subset \mathbb{R}$. Hence, it remains to show that the entropy integral $\int_0^\infty \sqrt{\log N(I_{n,\delta},\eta)}\,d\eta$ is finite and can be made arbitrarily small by the choice of δ.

As already mentioned, the process $\mathbf{C}_n$ is zero mean complex Gaussian, and with (3.28) we find that

$$\begin{aligned}\mathrm{E}|\mathbf{C}_n(s) - \mathbf{C}_n(t)|^2 &= \mathrm{E}\big((\mathbf{C}_n(s) - \mathbf{C}_n(t))\overline{(\mathbf{C}_n(s) - \mathbf{C}_n(t))}\big)\\ &= 2 - \widetilde{F}(s-t) - \widetilde{F}(t-s) - |\widetilde{F}(s) - \widetilde{F}(t)|^2\\ &\lesssim 1 - \mathrm{Re}\,\widetilde{F}(|s-t|)\,.\end{aligned}$$

By (3.20), there exists $\gamma > 0$ such that $y^\gamma F(-y) = O(1)$ and $y^\gamma(1-F(y)) = O(1)$ as $y \to \infty$. Without loss of generality, it can be assumed that $\gamma < 2$, so that from Theorem 11.3.2 in Kawata (1972) it follows that $1 - \mathrm{Re}\,\widetilde{F}(t) = O(t^\gamma)$ as $t \downarrow 0$. Conclude that

$$d(s,t) = O\big(|s-t|^{\gamma/2}\big) \quad \text{as } |s-t| \to 0\,.$$

As a result, there exists an absolute constant C such that $d(s,t) \le C|s-t|^{\gamma/2}$ for sufficiently small δ and $s,t \in I_{n,\delta}$, so that each $\eta^{2/\gamma}/C$-cover of $I_{n,\delta}$ with respect to the absolute value distance is an valid η-cover with respect to d, yielding

$$N(I_{n,\delta},\eta) \le 4C\delta\eta^{-2/\gamma}\,.$$

This and noting that $N(I_{n,\delta},\eta) = 1$ whenever $\eta \ge C(2\delta)^{\gamma/2} \ge \mathrm{diam}(I_{n,\delta})$ gives

$$\begin{aligned}\int_0^\infty \sqrt{\log N(I_{n,\delta},\eta)}\,d\eta &\le \int_0^{C(2\delta)^{\gamma/2}} \sqrt{\log 4C\delta\eta^{-2/\gamma}}\,d\eta\\ &= C(2\delta)^{\gamma/2}\int_0^1 \sqrt{\log 2C^{1-2/\gamma}x^{-2/\gamma}}\,dx\end{aligned}$$

which is $O(\delta^{\gamma/2})$ as $\delta \to 0$. □

Proof of Theorem 3.4. For all $\varepsilon > 0$ and $T_n \to \infty$, with $\mathbf{C}_{1,n}$ as given in (3.27),

$$\mathrm{P}\big(|\widetilde{\mathbf{F}}_n(h_n) - \mathbf{C}_{1,l_n}(h_n)| > \varepsilon\big) \le \mathrm{P}\Big(\sup_{0\le y\le T_n} |\widetilde{\mathbf{F}}_n(y) - \mathbf{C}_{1,l_n}(y)| > \varepsilon\Big) + \mathrm{P}(h_n > T_n)\,,$$

so that, choosing $T_n = 2t_n$ which is $o\big(n^{\gamma/(2\gamma+4)}(\log n)^{-(\gamma+1)/(\gamma+2)}\big)$ by (3.21), from Lemma 3.2 and (3.22) it follows that $\widetilde{\mathbf{F}}_n(h_n) = \mathbf{C}_{1,l_n}(h_n) + o_P(1)$. With this, Lemma 3.3 immediately shows that $\widetilde{\mathbf{F}}_n(h_n) = \mathbf{C}_{1,l_n}(t_n) + o_P(1)$, and all the same we find that $\widetilde{\mathbf{G}}_n(h_n) = \mathbf{C}_{2,m_n}(t_n) + o_P(1)$. Hence, in view of (3.16) and since $\bar{\eta}_n \to_p \eta$,

$$\sqrt{s_n}(\eta_n - \bar{\eta}_n) = \frac{\sqrt{s_n/m_n}}{\widetilde{G}_n(h_n)}\Big(\sqrt{\tau}\,\mathbf{C}_{1,l_n}(t_n) - \eta\mathbf{C}_{2,m_n}(t_n)\Big) + o_P(1)\,.$$

Below we show that

$$\frac{\sqrt{s_n/m_n}}{\widetilde{G}_n(h_n)} = (1 + o_P(1))\,\frac{|\widetilde{G}(t_n)|}{\widetilde{G}(t_n)}\,, \tag{3.30}$$

so that defining $z_n = |\widetilde{G}(t_n)|/\widetilde{G}(t_n)$,

$$\sqrt{s_n}\big(\eta_n - \bar{\eta}_n\big) = \frac{z_n}{(1+o_P(1))}\Big(\sqrt{\tau}\,\mathbf{C}_{1,l_n}(t_n) - \eta\,\mathbf{C}_{2,m_n}(t_n)\Big) + o_P(1)\,. \tag{3.31}$$

The conclusion follows by using (3.29), the fact that the $\mathcal{CN}(0,1,0)$-distribution is invariant under multiplication with the complex non-random numbers z_n with $|z_n| = 1$, and the independence of $\mathbf{C}_{1,n}$ and $\mathbf{C}_{2,n}$.

To show (3.30), from (3.29) it follows that $\widetilde{\mathbf{G}}_n(h_n) = O_P(1)$, which implies $|\widetilde{G}_n(h_n) - \widetilde{G}(h_n)| = o_P\big(\sqrt{s_n/m_n}\big)$. By (3.23) this gives

$$\big|\widetilde{G}_n(h_n) - \widetilde{G}(t_n)\big| = o_P\big(\sqrt{s_n/m_n}\big).$$

Applying (3.18) we therefore have

$$\begin{aligned}\frac{\widetilde{G}(t_n)}{\widetilde{G}_n(h_n)} &= \frac{\widetilde{G}_n(h_n) + o_P\big(\sqrt{s_n/m_n}\big)}{\widetilde{G}_n(h_n)} = 1 + \frac{|\widetilde{G}_n(h_n)|}{\widetilde{G}_n(h_n)}\,\frac{o_P\big(\sqrt{s_n/m_n}\big)}{\sqrt{s_n/m_n}(1+o_P(1))}\\ &= 1 + o_P(1)\,.\end{aligned}$$

Since $\big||\widetilde{G}_n(h_n)| - |\widetilde{G}(t_n)|\big| \le \big|\widetilde{G}_n(h_n) - \widetilde{G}(t_n)\big|$ by triangle inequality, we further have $|\widetilde{G}(t_n)| = \sqrt{s_n/m_n}(1+o_P(1))$, and thus

$$\frac{\sqrt{s_n/m_n}}{\widetilde{G}_n(h_n)} = \frac{\widetilde{G}(t_n)}{\widetilde{G}_n(h_n)}\,\frac{\sqrt{s_n/m_n}}{\widetilde{G}(t_n)} = (1+o_P(1))\,\frac{|\widetilde{G}(t_n)|}{\widetilde{G}(t_n)}\,.$$

□

The proof of Theorem 3.5 is given in Hohmann and Holzmann (2013c).

Proofs of Section 3.2

Proof of Theorem 3.1. From C1, for $i = 0,1$

$$\frac{F(y|B_i)}{(1-\pi(B_i))\,F_0(y)} = 1 + \frac{\pi(B_i)\,F_1(y)}{(1-\pi(B_i))\,F_0(y)} \to 1, \qquad y \to -\infty,$$

and similarly under C2 and C3

$$\begin{aligned}1 - F(y|B_i) &\sim \pi(B_i)\,(1 - F_1(y)) && \text{as } y \to +\infty,\\ \widetilde{F}(y|B_i) &\sim \pi(B_i)\,\widetilde{F}_1(y) && \text{as } y \to +\infty,\end{aligned}$$

Therefore, as $y \to -\infty$,

$$\frac{F(y|B_1)}{F(y|B_0)} \sim \frac{(1-\pi(B_1))F_0(y)}{(1-\pi(B_1))F_0(y)} = \frac{1-\pi(B_1)}{1-\pi(B_1)} = \zeta,$$

which is (3.6), and (3.7) follows similarly. Then, given the identification of ζ and ξ, one determines the component cdfs F_0 and F_1 by applying (3.4), and the mixture weight π by using

(3.5).

Now let $G(y|B) = (1-\mu(B))G_0(y) + \mu(B)G_1(y)$ be another mixture in $\mathscr{M}_i$ such that $G = F$. Then, by C1,

$$\zeta = \lim_{y\to-\infty} \frac{F(y|B_1)}{F(y|B_0)} = \lim_{y\to-\infty} \frac{G(y|B_1)}{G(y|B_0)}$$
$$= \lim_{y\to-\infty} \frac{1+\mu(B_1)\big(G_1(y)/G_0(y)-1\big)}{1+\mu(B_0)\big(G_1(y)/G_0(y)-1\big)} = \frac{1-\mu(B_1)}{1-\mu(B_0)}\,,$$

from which we conclude that $\mu(B_0) < 1$ and $\mu(B_0) \neq \mu(B_1)$ since $\zeta \neq 1$. Similarly, we obtain $\mu(B_0) > 0$ by C2 respectively C3. Thus, (3.4) applies to G as well, and since $F = G$ and the values for ξ and ζ coincide, (3.4) implies that $F_i = G_i$, $i = 0,1$. In particular, $G_0(y_0) \neq G_1(y_0)$, and by (3.5) it follows that $\pi = \mu$. □

Proofs of Section 3.3

Proof of Proposition 3.1. Using our previous notation, set $\bar{\zeta}_n = F(L_n|B_1)/F(L_n|B_0)$ and $\bar{\xi}_n = \mathrm{Re}(\widetilde{F}(R_n|B_1)/\widetilde{F}(R_n|B_0))$. Let $\mathscr{Z} = \sigma(\mathbf{1}_{\{Z_k\in B_j\}}, k\in\mathbb{N}, j=0,1)$ be the σ-field generated by the indicator variables $\mathbf{1}_{\{Z_k\in B_j\}}$. By A 3.2, the distribution of $Y_1,\dots,Y_n$ conditional on $\mathscr{Z}$ is given by

$$\mathrm{P}(Y_1\le y_1,\dots,Y_n\le y_n|\mathscr{Z})$$
$$= \prod_{\substack{1\le i\le n\\ Z_i\in B_0}} F(y_i|B_0) \prod_{\substack{1\le i\le n\\ Z_i\in B_1}} F(y_i|B_1) \prod_{\substack{1\le i\le n\\ Z_i\notin B_0\cup B_1}} F(y_i|(B_0\cup B_1)^{\complement}),$$

and thus, evaluating the latter at a fixed ω,

$$\mathrm{P}(Y_i^*\le y_i^*, Y_j^{**}\le y_j^{**}, i=1,\dots,m_n, j=1,\dots,l_n|\mathscr{Z})(\omega)$$
$$= \prod_{i=1}^{m_n(\omega)} F(y_i^*|B_0) \prod_{j=1}^{l_n(\omega)} F(y_j^{**}|B_1).$$

Hence, conditional on $\mathscr{Z}$, the variables $Y_1^*,\dots,Y_{m_n}^*$ and $Y_1^{**},\dots,Y_{l_n}^{**}$ are independent i.i.d. samples from $F(\cdot|B_0)$ and $F(\cdot|B_1)$, resp., where for almost all ω it holds that

$$m_n(\omega)/l_n(\omega) \to p_0/p_1\,.$$

Defining $\tau = p_0/p_1$, it therefore follows from Theorem 3.5 and (3.12) that

$$\sqrt{r_n}\begin{pmatrix}\zeta_n-\zeta\\ \xi_n-\xi\end{pmatrix}\Big|\mathscr{Z} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix}\beta_\zeta\\ \beta_\xi\end{pmatrix}, \begin{pmatrix}\tau\zeta+\zeta^2 & 0\\ 0 & (\tau+\xi^2)/2\end{pmatrix}\right)$$

with probability one. Finally, since the weak limit does actually not depend on the specific realization of the sequence $\mathbf{1}_{\{Z_k\in B_j\}}$, the result follows. □

Proof of Theorem 3.2. We only consider the process $\mathbf{F}_n = \sqrt{r_n}\big(\hat{F}_1 - F_1\big)$. Let $K(x) = (1-x)^{-1}$,

$x \neq 1$. Since

$$\hat{F}_1(y) - F_1(y) = F_n(y|B_0) - F(y|B_0) + \big(K(\zeta_n) - K(\zeta)\big)\, D(F_n, y) \\ + \big(D(F_n, y) - D(F, y)\big)\, K(\zeta), \tag{3.32}$$

where $F_n(y|B_i) - F(y|B_i) = O_P\big(n^{-1/2}\big)$ and $r_n/n \to 0$, with $\mathrm{K}_n = \sqrt{r_n}(K(\zeta_n) - K(\zeta))$ we find that

$$\mathbf{F}_n(y) = D(F_n, y)\mathrm{K}_n + o_P(1).$$

The function K being differentiable at ζ with derivative $(1-\zeta)^{-2}$, the Delta method and (3.13) yield

$$\mathrm{K}_n \rightsquigarrow \mathcal{N}\big((1-\zeta)^{-2}\beta_\zeta, (1-\zeta)^{-4}\sigma_\zeta^2\big).$$

Since $D(F_n, y) \to_{a.s.} D(F, y)$, we therefore find that, for any $y_1, y_2 \in \mathbb{R}$, defining $\mathrm{D}(y_1, y_2) = \big(D(F, y_1), D(F, y_2)\big)'$,

$$\big(\mathbf{F}_n(y_1), \mathbf{F}_n(y_2)\big)' \rightsquigarrow \mathcal{N}\big(\mathrm{D}(y_1,y_2)(1-\zeta)^{-2}\beta_\zeta, \mathrm{D}(y_1,y_2)\mathrm{D}(y_1,y_2)'(1-\zeta)^{-4}\sigma_\zeta^2\big),$$

which shows the weak convergence of the finite dimensional distributions of $\mathbf{F}_n$ to those of the Gaussian process $\mathbf{G}_\zeta$.

To conclude weak convergence in $\ell^\infty(\mathbb{R})$, it remains to show that $\mathbf{F}_n$ is asymptotically tight, a sufficient condition for which is to show that for any $\varepsilon, \eta > 0$ we can find a finite partition $\{I_1, \ldots, I_d\}$ of $\mathbb{R}$ such that

$$\limsup_{n\to\infty} \mathrm{P}\Big(\max_{k=1,\ldots,d} \sup_{s,t \in I_k} \big|\mathbf{F}_n(t) - \mathbf{F}_n(s)\big| > \varepsilon\Big) < \eta \tag{3.33}$$

holds, cf. van der Vaart and Wellner (2000, Theorem 1.5.6). Similar to (3.32), write

$$\mathbf{F}_n(s) - \mathbf{F}_n(t) = \big(F_n(s|B_1) - F_n(t|B_1) + F_n(t|B_0) - F_n(s|B_0)\big)\, \mathrm{K}_n + o_P(1),$$

where the remainder is uniform in s, t. By the uniform convergence of F_n,

$$\sup_{s,t\in\mathbb{R}} \big|F_n(s|B_i) - F_n(t|B_i)\big| \le \sup_{s,t\in\mathbb{R}} \big|F(s|B_i) - F(t|B_i)\big| + o_P(1).$$

For all $\gamma \in (0,1)$ one can find a finite partition $\{I_k\}$ of $\mathbb{R}$ such that

$$\sup_{s,t\in I_k} \big(\big|F(s|B_1) - F(t|B_1)\big| + \big|F(s|B_0) - F(t|B_0)\big|\big) < \gamma,$$

implying

$$\max_k \sup_{s,t\in I_k} \big|\mathbf{F}_n(s) - \mathbf{F}_n(t)\big| \le \gamma|\mathrm{K}_n| + o_P(1).$$

By asymptotic normality of K_n, we find $C > 0$ such that

$$\limsup_n \mathrm{P}\big(|\mathrm{K}_n| > C\big) < \eta.$$

Now choosing $\gamma = C/\varepsilon$ gives (3.33) and hence the asymptotic tightness of $\mathbf{F}_n$. □

Theorem 3.3 is proved by using the Delta method.

4

Weighted angle Radon transform: Convergence rates and efficient estimation

Daniel Hohmann and Hajo Holzmann

Abstract: In the statistics literature, recovering a signal which is observed under the Radon transform is considered as a very mildly ill-posed inverse problem. In this paper, we argue that several statistical models which involve the Radon transform lead to an observational design which strongly influences its degree of ill-posedness, and that the Radon transform may actually become severely ill-posed. The main ingredient here is a weight function λ on the angle. Extending results for the limited angle situation, we compute the singular value decomposition of the Radon transform as an operator between suitably weighted L_2-spaces, and show how the singular values relate to the eigenvalues of the sequence of Toeplitz matrices of λ. Further, in the associated white noise sequence model, we give upper and lower bounds on the rate of convergence, and in several special situations even obtain optimal rates with precise minimax constants. For the severely ill-posed limited angle problem, a simple projection estimator is adaptive in the exact minimax sense.

Keywords: nonparametric estimation, Radon transform, limited angle problem, efficient estimation, minimax estimation

MSC2010 classification: 62G10, 62G20

4.1. Introduction

Recovering images (functions) observed under the Radon transform is one of the most important and common inverse problems, with fundamental applications in tomography and other fields, see e.g. Natterer (1986) for an overview. In the statistics literature, which has devoted a significant amount of effort to the issue (see below for a review of the literature), this inverse problem is considered to be only very mildly ill-posed. In this paper, however, we show that the ill-posedness of the Radon transform strongly depends on the observational design, and that observational designs which lead to significantly more severe ill-posedness arise naturally in statistical models involving the Radon transform. We shall restrict attention to the two-dimensional case, in which The Radon transform is said to be only mildly ill-posed of degree $1/2$.

Let $B_1(0) = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$ be the unit disc in $\mathbb{R}^2$ and let $f : B_1(0) \to \mathbb{R}$ be integrable. Then its Radon transform is defined (for almost all (φ, s)) as

$$\mathrm{R}f(\varphi,s) = \int_{|t|\leq\sqrt{1-s^2}} f(s\cos\varphi - t\sin\varphi, s\sin\varphi + t\cos\varphi)\,dt,$$
$$(\varphi,s) \in [-\pi/2,\pi/2]\times[-1,1].$$

We shall follow Johnstone and Silverman (1990) and call the domain $[-\pi/2,\pi/2]\times[-1,1]$ of $\mathrm{R}f$ the detector space, and $B_1(0)$ brain space. The aim is to estimate f from noisy data on its Radon transform.

We shall argue that due to the observational design, the Radon transform needs to be studied as an operator between weighted L_2-spaces

$$\begin{aligned} &\mathrm{R} : L_2(B_1(0);\mu_2) \longrightarrow L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1), \\ &d\mu_2(x,y) = w_2(x,y)\,dx\,dy, \quad d\mu_1(\varphi,s) = \lambda(\varphi) w_1(s)\,d\varphi\,ds. \end{aligned} \tag{4.1}$$

Here, the most striking feature is the weight function $\lambda : [-\pi/2,\pi/2] \to [0,\infty)$ on the angle in detector space. The case when λ has support $[-\eta,\eta]$ for some $\eta < \pi/2$ is called the limited angle Radon transform (cf. Davison, 1983). However, it turns out that even if λ only has two zeros at the boundary points $\pm\pi/2$, the degree of ill-posedness of R will depend on λ. For the weight functions w_1 and w_2, we consider the following parametric families in $\gamma > -1/2$,

$$\begin{aligned} w_1(s) &= \frac{\sqrt{\pi}\,\Gamma(\gamma+1/2)}{\gamma\Gamma(\gamma)}(1-s^2)^{1/2-\gamma}, \qquad -1 \leq s \leq 1, \\ w_2(x,y) &= \frac{\pi}{\gamma}(1-x^2-y^2)^{1-\gamma}, \qquad (x,y) \in B_1(0). \end{aligned}$$

The weight function w_1 in detector space also corresponds to the measurement design, the most important cases being $\gamma = 1$ (fan beam design) as well as $\gamma = 1/2$ (parallel beam design). The weight w_2 with corresponding γ is then required for technical reasons, in order to make the singular value decomposition (SVD) of R analytically tractable. In particular, in the parallel beam design $\gamma = 1/2$, the estimation error in brain space is measured with a weighted L_2-norm.

Statistical models

We discuss statistical models which involve the weight function λ, and also indicate the appropriate values of the parameter γ in w_1 and w_2.

1. Nonparametric random coefficient regression models. Nonparametric estimation in random coefficient regression models was first studied in Beran, Feuerverger and Hall (1996) and Feuerverger and Hall (1996). These models have recently become quite popular in econometrics, see Hoderlein, Klemelä and Mammen (2010) and Gautier and Kitamura (2013). Suppose that we observe (Y,X) from the model $Y = X^T\beta$. Here $X,\beta \in \mathbb{R}^2$ are independent random vectors, and the unobserved β has a Lebesgue density f_β supported in $B_1(0)$. The aim is to estimate f_β. If we standardize $Z = Y/\|X\|$, $X/\|X\| = (\cos(\Phi),\sin(\Phi))$, then

$$f_{Z|\Phi=\varphi}(z) = (\mathrm{R}f_\beta)(\varphi,z).$$

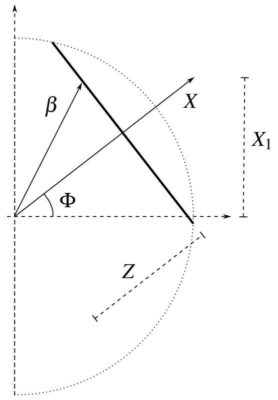


Figure 4.1: Parametrization in the random coefficient model. The bold line is the set of all $\beta \in B_1(0)$ for which $\beta' X/\|X\| = Z$.

Given $h(\varphi,s)$, if Φ has a Lebesgue density f_Φ we have

$$\begin{aligned} \mathrm{E}\big(h(\Phi,Z)\big) &= \int_{-\pi/2}^{\pi/2}\int_{-1}^{1} h(\varphi,z) f_{Z|\Phi=\varphi}(z)\, f_\Phi(\varphi)\, dz\, d\varphi \\ &= \int_{-\pi/2}^{\pi/2}\int_{-1}^{1} (\mathrm{R} f_\beta)(\varphi,z)\, h(\varphi,z)\, d\mu(\varphi,z), \end{aligned}$$

where $d\mu(\varphi,z) = f_\Phi(\varphi) dz d\varphi = 2^{-1} d\mu_1(\varphi,z)$ with $\lambda(\varphi) = f_\Phi(\varphi)$ and $\gamma = 1/2$. If $X = (1,X_1)$ includes an intercept as well as an additional covariate, the support of X_1 will determine the support of Φ, and in case of full support of X_1 with density f_{X_1}, the tails of X_1 determine the rate of decay of f_Φ at $\pm\pi/2$ since $f_\Phi(\varphi) = f_{X_1}(\tan\varphi)(1+(\tan\varphi)^2)$. Thus, only for quite heavy tails of X_1 (Cauchy-type tails) is f_Φ bounded away from 0, which is the case studied in Hoderlein et al. (2010). See Figure 4.1 for an illustration. Our results will show that for lighter tails, the Radon transform R on the weighted L_2-spaces is in fact more ill-posed.

2. Regression. Suppose that we observe random variables (Y,Θ,S) from the model

$$Y = (\mathrm{R}f)(\Theta,S) + \varepsilon, \qquad \mathrm{E}(\varepsilon|\Theta,S) = 0.$$

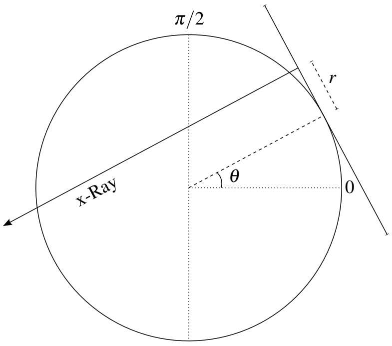
If (Θ,S) is distributed according to μ_1, then given $h(\varphi,s)$,

$$\mathrm{E}\big(Y h(\Theta,S)\big) = \langle (\mathrm{R}f), h\rangle_{\mu_1}.$$

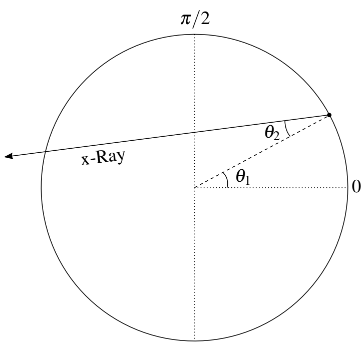
This is a statistical framework for computerized tomography (Natterer, 1986), and the measure μ_1 is determined by the measurement design. The case $\gamma = 1$ corresponds to the fan beam design, the case $\gamma = 1/2$ to the parallel beam design, see Figure 4.2.

For the fan beam design, most statistical literature uses SVD based or derived methods (such as needlets), see Cavalier and Tsybakov (2002); Kerkyacharian, Kyriazis, le Pennec, Petrushev and Picard (2010); Kerkyacharian, le Pennec and Picard (2012). In case of parallel beam, Korostelev and Tsybakov (1993) as well as Cavalier (1998) use estimates based on the filtered back-projection algorithm.

No paper in the statistics literature seems to take into account a weight function λ on the angle,



(a) Parallel beam: $\gamma = 1/2$



(b) Fan beam: $\gamma = 1$

Figure 4.2: Parametrization of the measurement design in computerized tomography: Measurements are performed uniformly distributed on (a) $[-\pi/2,\pi/2]\times[-1,1]$ in parallel beam design and (b) $[-\pi/2,\pi/2]^2$ in fan beam design.

which arises most naturally in the parallel beam design in form of a limited angle. The bias of the filtered back-projection algorithm in case of limited angle is discussed in Frikel (2013). We shall derive SVD-based methods which also work for parallel beam design and take care of the weight function λ.

3. Density estimation. Johnstone and Silverman (1990) propose a model of Positron emission tomography in which the emission density $f(x_1,x_2)$ on $B_1(0)$ needs to be estimated from data (Θ,S) distributed according to Rf. Here $\mathrm{E}\big(g(\Theta,S)\big) = \langle(\mathrm{R}f),g\rangle_{\mu_1}$ without weight functions ($\gamma = 1/2$ and $\lambda = 1$). In order to take advantage of the simpler form of the singular value decomposition in case $\gamma = 1$, they insert the weight w_1 with $\gamma = 1$ into $\mathrm{E}\big(g(\Theta,S)w_1(\Theta)\big)$. As a consequence, the variance term in the risk is difficult to handle, and therefore they resort to a surrogate mean integrated squared error in order to measure the precision of their estimators.

They also discuss missing data problems, in which certain photons may not be observed at random. In this case, the actual observations have density proportional to $\mathrm{R}f(\varphi,s)a(\varphi,s)$ for some function $0\le a(\varphi,s)\le 1$. Johnstone and Silverman (1990) show that the minimax rates of convergence remain unchanged if $a(\varphi,s)$ is bounded away from zero. Our results indicate that without this additional assumption, the rate of convergence can become even logarithmically slow.

We shall conduct our convergence analysis in the idealized white noise model. For direct regression as well as density estimation problems, asymptotic equivalence to a white noise model has been obtained in Brown and Low (1996), Nussbaum (1996), and Reiß (2008). While no corresponding results are available for the above indirect models yet, the analysis in the technically less complicated white noise model still gives a valuable insight into the difficulty of the estimation problem.

The paper is structured as follows. In Section 4.2 we derive the singular value decomposition of the Radon transform between the weighted L_2-spaces as in (4.1). The singular values relate to the eigenvalues of the sequence of Toeplitz matrices defined by λ, and we present the relevant results from the literature on their asymptotic behavior. In Section 4.3 we study efficient estimation in a white noise sequence model. In Section 4.3.2 we consider the severely ill-posed limited angle problem, in which a simple projection estimator is even sharp minimax adaptive. In Section 4.3.3

we give upper and lower bounds on the rate of convergence in case of polynomial decay of the singular values, and in Section 4.3.4 we obtain precise rates with asymptotic minimax constants for the fan beam design ($\gamma = 1$) and for a class of weight functions λ with banded Toeplitz matrix. Section 4.4 concludes, while proofs are deferred to Section 4.5. Some further technical results can be found in the supplementary Appendix.

4.2. Singular value decomposition

Davison (1983) presents the SVD of the Radon transform with weight functions w_1 and w_2, without weight on the angle. Further, in case of limited angle and $\gamma = 1$, he relates the singular values to the eigenvalues of certain hermitian Toeplitz matrices.

We extend his analysis by allowing a general weight function λ on the angle as well as general parameter $\gamma > -1/2$ for the weighted Radon transform R in (4.1), and also present explicitly the singular functions involved.

To start, in the supplement, Lemma 4.7, we show that if λ is integrable, then R is continuous with norm

$$\|\mathrm{R}\|^2 = \sup_{\|f\|_{\mu_2}=1} \|\mathrm{R}f\|^2_{\mu_1} = \int_{-\pi/2}^{\pi/2} \lambda(\varphi)\,d\varphi.$$

Set

$$\phi_m = w_1^{-1} C_m^\gamma, \qquad m = 0, 1, \dots,$$

with C_m^γ the Gegenbauer or ultraspherical polynomials on $[-1,1]$. The ϕ_m are orthogonal and complete in $L_2([-1,1]; w_1(s)\,ds)$. Next, Lemma 4.9 in the supplement shows that for a function $g(\varphi, s) = h(\varphi)\phi_m(s)$ with h integrable on $[-\pi/2, \pi/2]$,

$$(\mathrm{RR}^* g)(\varphi, s) = \frac{\phi_m(s)}{C_m^\gamma(1)} \int_{-\pi/2}^{\pi/2} h(\varphi') C_m^\gamma(\cos(\varphi' - \varphi)) \lambda(\varphi')\,d\varphi',$$

where R^* is the adjoint of the Radon transform. This shows that RR^* leaves the subspaces V_m of $L_2([-\pi/2, \pi/2] \times [-1,1]; \mu_1)$ consisting of all functions of the form $g = h\phi_m$ invariant, where in particular these subspaces V_m span the whole of $L_2([-\pi/2, \pi/2] \times [-1,1]; \mu_1)$ by completeness of the ϕ_m. Hence, it remains to study the action of the self-adjoint integral operators T_m on $L_2([-\pi/2, \pi/2], \lambda(\varphi)d\varphi)$ given by

$$T_m h(\varphi) = C_m^\gamma(1)^{-1} \int_{-\pi/2}^{\pi/2} h(\varphi') C_m^\gamma(\cos(\varphi' - \varphi)) \lambda(\varphi')\,d\varphi'.$$

In the supplement, Lemma 4.10, we further show that T_m vanishes on the orthogonal complement of $\mathrm{lin}\{h_{m,k}\}_{k=0}^m$ with $h_{m,k}(\varphi) = e^{-i(m-2k)\varphi}$, and that its action on $\mathrm{lin}\{h_{m,k}\}_{k=0}^m$ is determined as follows.

Writing $\mathrm{h}_m = (h_{m,0}, \dots, h_{m,m})'$ and $T_m \mathrm{h}_m = (T_m h_{m,0}, \dots, T_m h_{m,m})'$, we have

$$T_m \mathrm{h}_m = \pi(\mathrm{C}_m \mathrm{A}_m)' \mathrm{h}_m, \tag{4.2}$$

where

$$\mathrm{C}_m = \mathrm{diag}(c_{m,0}, \dots, c_{m,m}), \qquad c_{m,j} = \binom{m}{j} \frac{\Gamma(2\gamma)\Gamma(j+\gamma)\Gamma(m-j+\gamma)}{\Gamma(m+2\gamma)\Gamma(\gamma)^2}, \tag{4.3}$$

and

$$\mathrm{A}_m = \big(a_{j-k}\big)_{j,k=0,\ldots,m}, \qquad m = 0,1,2,\ldots, \tag{4.4}$$

is the Toeplitz matrix determined by the sequence

$$a_z = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-iz\varphi'}\,\lambda(\varphi'/2)\,d\varphi' = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} e^{-i2z\varphi'}\,\lambda(\varphi')\,d\varphi', \qquad z\in\mathbb{Z}.$$

Even though T_m is self-adjoint, the matrix $\mathrm{C}_m\mathrm{A}_m$ is not Hermitian (since $h_{m,j}$, $j = 1,\ldots,m$, are not orthonormal in $L_2([-\pi/2,\pi/2];\lambda(\varphi)d\varphi)$. To find an orthonormal system of eigenfunctions of T_m that span the range of T_m, and hence to complete the SVD of R, we first require a basis of $\mathrm{lin}\{h_{m,k}\}_{k=0}^m$ which is orthonormal in $L_2([-\pi/2,\pi/2];\lambda(\varphi)\,d\varphi)$.

To this end, note that A_m is Hermitian so that there exists an orthonormal system $\{\mathrm{v}_{m,l} = (v_{m,l}^{(0)},\ldots,v_{m,l}^{(m)})'\}_{l=0}^m$ of eigenvectors of A_m with corresponding real eigenvalues $\alpha_{m,0}\geq\ldots\geq\alpha_{m,m}>0$. It is well known that the $\alpha_{m,l}$ are strictly positive whenever λ is not essentially zero (which we shall always assume), see for instance Tilli (2003) for universal lower bounds on the smallest eigenvalues of sequences of Toeplitz matrices. Then

$$\tilde{h}_{m,l} = \frac{1}{\sqrt{\pi\alpha_{m,l}}}\mathrm{v}'_{m,l}\mathrm{h}_m = \frac{1}{\sqrt{\pi\alpha_{m,l}}}\sum_{k=0}^m v_{m,l}^{(k)}h_{m,k}, \qquad l = 0,\ldots,m,$$

is an orthonormal basis of $\mathrm{lin}\{h_{m,k}\}_{k=0}^m$ in $L_2([-\pi/2,\pi/2];\lambda(\varphi)\,d\varphi)$, and again setting $\tilde{\mathrm{h}}_m = (\tilde{h}_{m,0},\ldots,\tilde{h}_{m,m})'$, we have that

$$T_m\tilde{\mathrm{h}}_m = \mathrm{B}'_m\tilde{\mathrm{h}}_m, \qquad \mathrm{B}_m = \pi\Lambda_m^{1/2}\mathrm{V}_m^*\mathrm{C}_m\mathrm{V}_m\Lambda_m^{1/2}.$$

Choose an orthonormal system $\{\mathrm{w}_{m,l} = (w_{m,l}^{(0)},\ldots,w_{m,l}^{(m)})'\}_{l=0}^m$ of eigenvectors of $\pi^{-1}\mathrm{B}_m$ with corresponding eigenvalues $\beta_{m,0},\ldots,\beta_{m,m}$ (the same as those of $\mathrm{C}_m\mathrm{A}_m$), and set

$$\tilde{\tilde{h}}_{m,l} = \mathrm{w}'_{m,l}\tilde{\mathrm{h}}_m = \sum_{k_1,k_2=0}^m \frac{w_{m,l}^{(k_1)}v_{m,k_1}^{(k_2)}}{\sqrt{\pi\alpha_{m,k_1}}}h_{m,k_2}. \tag{4.5}$$

It readily follows that $\tilde{\tilde{h}}_{m,0},\ldots,\tilde{\tilde{h}}_{m,m}$ are orthonormal in $L_2([-\pi/2,\pi/2];\lambda(\varphi)\,d\varphi)$, and by definition of B_m we have $T_m\tilde{\tilde{h}}_{m,l} = \pi\beta_{m,l}\tilde{\tilde{h}}_{m,l}$. Also the $\beta_{m,l}$ are strictly positive since the eigenvalues of both A_m and C_m are.

Set $\tilde{\phi}_m = d_m^{-1/2}\phi_m$, where

$$d_m = \langle\phi_m,\phi_m\rangle_{w_1} = \frac{\sqrt{\pi}\gamma 2^{1-2\gamma}\Gamma(m+2\gamma)}{m!(m+\gamma)\Gamma(\gamma)\Gamma(\gamma+1/2)},$$

and

$$\Phi_{m,l}(\varphi,s) = \tilde{\phi}_m(s)\tilde{\tilde{h}}_{m,l}(\varphi), \qquad -\pi/2\leq\varphi\leq\pi/2,\ -1\leq s\leq 1. \tag{4.6}$$

Altogether we found that, for all $m\geq l\geq 0$,

$$\mathrm{R}\mathrm{R}^*\Phi_{m,l} = \tilde{\phi}_m T_m\tilde{\tilde{h}}_{m,l} = \pi\beta_{m,l}\Phi_{m,l}, \tag{4.7}$$

where the system $\{\Phi_{m,l}\}_{m\geq l\geq 0}$ is orthonormal in $L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1)$ and complete in range(R), so we finally arrived at

Theorem 4.1. *For all $m\geq l\geq 0$, let $\Phi_{m,l}$ be defined as in (4.6), and $\beta_{m,0},\dots,\beta_{m,m}$ be the eigenvalues of the matrix product $\mathrm{C}_m\mathrm{A}_m$ with C_m and A_m defined in (4.3) and (4.4), respectively. Set $\sigma_{m,l}=\sqrt{\pi\beta_{m,l}}$ and $\Psi_{m,l}=\sigma_{m,l}^{-1}\mathrm{R}^*\Phi_{m,l}$. The singular value decomposition of* R *between the weighted L_2-spaces in (4.1) is then given by*

$$\left\{\Psi_{m,l},\Phi_{m,l},\sigma_{m,l}\right\}_{m\geq l\geq 0}.$$

In particular, the functions $(\Psi_{m,l})_{m\geq l\geq 0}$ form an orthonormal basis of $L_2(B_1(0);\mu_2)$, so that R *is injective, and we have for all $f\in L_2(B_1(0);\mu_2)$ that*

$$f=\sum_{m=0}^{\infty}\sum_{l=0}^{m}\sigma_{m,l}^{-1}\langle \mathrm{R}f,\Phi_{m,l}\rangle_{\mu_1}\Psi_{m,l}.$$

To complete the proof of the theorem, in Lemma 4.11 in the appendix we determine the functions $\Psi_{m,l}=\sigma_{m,l}^{-1}\mathrm{R}^*\Phi_{m,l}$ explicitly, and show that they form an orthonormal basis of $L_2(B_1(0);\mu_2)$.

The case $\gamma=1$: Fan beam design. Let us specialyze our results for the case $\gamma=1$. Here the weights $c_{m,l}$ have the simple form $c_{m,l}=(m+1)^{-1}$ for all m, so, given the eigenvalues $\alpha_{m,l}$ of A_m, it follows that $\beta_{m,l}=\alpha_{m,l}/(m+1)$, and thus the singular values of the operator R are

$$\sigma_{m,l}=\sqrt{\frac{\pi\alpha_{m,l}}{m+1}},\qquad m\geq l\geq 0. \tag{4.8}$$

The singular functions also simplify, see the supplement for explicit formulas.

Finally, following Johnstone (1989) we relate ellipsoid-type smoothness conditions to certain weak derivatives w.r.t. a weighted L_2-norm. To this end, introduce the measure

$$d\mu_3(x,y)=\pi^{-1}(s+1)(1-x^2-y^2)^s\,dx\,dy,\qquad (x,y)\in B_1(0).$$

Proposition 4.1. *A function $f\in L_2(B_1(0);\mu_2)$ has weak derivatives of order s in the weighted L_2-space $L_2(B_1(0);\mu_3)$ if and only if its Fourier coefficients $\theta_{m,l}=\langle f,\Psi_{m,l}\rangle$, with singular base functions $\Psi_{m,l}$ given in (4.47), satisfy*

$$\sum_{m=0}^{\infty}\sum_{l,k=0}^{m}\theta_{m,l}^2\big(v_{m,l}^{(k)}\big)^2(m-k+1)^s(k+1)^s<\infty.$$

4.2.1. Asymptotics of the singular values

The ill-posedness of the weighted angle Radon transform is determined by the decay of the singular values

$$\sigma_{m,l}=\sqrt{\pi\beta_{m,l}}, \tag{4.9}$$

i.e. by the eigenvalues $\beta_{m,l}$ of the matrix product $\mathrm{C}_m\mathrm{A}_m$ with C_m and A_m defined in (4.3) and (4.4), respectively.

Decay of the $c_{m,l}$. In case $\lambda = 1$, the matrix A_m reduces to the identity matrix, and we have $\beta_{m,l} = c_{m,l}$. Using $\Gamma(x+\delta)/\Gamma(x) \sim x^{\delta}$ as $x \to \infty$ for all $\delta \in \mathbb{R}$, it is easily seen that the inner weights, those for which l grows as αm for some $\alpha \in (0,1)$, decay according to

$$c_{m,l} \sim \frac{\Gamma(2\gamma)}{\Gamma(\gamma)^2}(\alpha(1-\alpha))^{\gamma-1}(m+1)^{-1},$$

while the outer weights with l (or $m-l$) fixed behave like

$$c_{m,l} \sim \frac{\Gamma(2\gamma)}{\Gamma(\gamma)^2}\frac{\Gamma(l+\gamma)}{\Gamma(l+1)}(m+1)^{-\gamma},$$

both as $m \to \infty$. In particular, for $\gamma \leq 1$, the extreme weights satisfy

$$\min_{l=0,\dots,m} c_{m,l} \sim \frac{\Gamma(2\gamma)}{\Gamma(\gamma)^2 4^{\gamma-1}}(m+1)^{-1}, \qquad \max_{l=0,\dots,m} c_{m,l} \sim \frac{\Gamma(2\gamma)}{\Gamma(\gamma)}(m+1)^{-\gamma}. \tag{4.10}$$

For $\gamma > 1$ the roles of min and max are reversed.

Eigenvalues of the Toeplitz matrices A_m. We call A_m in (4.4) the Toeplitz matrix generated by the function λ. The asymptotic behavior of the eigenvalues of such sequences of Toeplitz matrices has been intensively studied in the literature. A famous result by Szegö, see Grenander and Szegö (1958), states that the averages of the eigenvalues of A_m tend to the normalized integral of $\lambda(\cdot/2)$. Further results mainly concern the extreme eigenvalues. We shall summarize results that we shall require below. Recall that the ordered eigenvalues of A_m are denoted by $\alpha_{m,0} \geq \ldots \geq \alpha_{m,m} > 0$.

For the weight function $\lambda = \mathbf{1}_{[-\eta,\eta]}$, where $\eta < \pi/2$, the Toeplitz matrices A_m generated by λ are given by

$$A_m = \left(\frac{\sin(2(j-k)\eta)}{\pi(j-k)}\right)_{j,k=0,\dots,m},$$

where for $j = k$ this expression is understood as the continuous continuation with value $2\eta/\pi$. It is well known that in this case the small eigenvalues of A_m decay to zero exponentially fast, see Slepian (1978), and specifically that

$$\alpha_{m,m} \sim C m^{1/2} e^{-\xi m} \qquad \text{as } m \to \infty, \tag{4.11}$$

where the constants $C, \xi > 0$ only depend on the angle η, and where ξ is given by

$$\xi = \log\left(1 + \frac{2\sqrt{1-\cos(\pi-2\eta)}}{\sqrt{2} - \sqrt{1-\cos(\pi-2\eta)}}\right). \tag{4.12}$$

Slepian (1978) also discusses the behaviour of the other extreme as well as of the intermediate eigenvalues, which we shall not require, however.

In case of a single root of λ (mod π, typically $\pi/2$), the extreme eigenvalues α_{mm} decay polynomially, with degree depending on the order of the root. More precisely, if $\lambda : \mathbb{R} \to \mathbb{R}_+$ is continuous and π-periodic, if there is a unique value φ_0 (mod π) such that $\lambda(\varphi_0) = 0$, and if there exists $\rho > 0$ such that, with $k = k(\rho) = \lfloor \rho/2 \rfloor$, $g(\varphi) = \lambda(\varphi)^{2k/\rho}$ has $2k$ continuous derivatives in some neighborhood of φ_0, and $g^{(2k)}$ is the first non-vanishing derivative of g at φ_0, then there exists $C > 0$ such that $\alpha_{m,m}^{-1} \sim Cm^{\rho}$, see Parter (1961). For example, for $\lambda = \cos^2$, $\alpha_{m,m}^{-1} \asymp m^2$.

In case of polynomial decay, the behaviour of the extreme eigenvalues of the matrices A_m alone does not suffice to determine the precise rate of convergence even in the fan beam design, uniform results are required. Böttcher, Grudsky and Maksimenko (2010a) studied the uniform behavior of the eigenvalues of banded Toeplitz matrices A_m, here we briefly review their main result. Let a be a Laurent polynomial

$$a(t) = \sum_{k=-r}^{r} a_k t^k, \qquad r \in \mathbb{N}, \quad a_r \neq 0, \quad \bar{a}_k = a_{-k},$$

and assume that the π-periodic weight function λ is given by $\lambda(\varphi) = a(e^{i2\varphi})$. In this case we say that λ is banded since, by construction, the Hermitian Toeplitz matrices A_m generated by λ are banded. In fact, the coefficients a_k of the polynomial a are exactly the entries of A_m. In particular, the condition $\bar{a}_k = a_{-k}$ ensures that λ is real.

Assumption A 4.1. The banded weight function λ satisfies $\lambda(-\pi/2) = \lambda(\pi/2) = 0$, there is a unique maximizer φ_0 such that λ is strictly increasing on $(-\pi/2, \varphi_0)$ and strictly decreasing on $(\varphi_0, \pi/2)$, and the second derivatives of λ at $\pi/2$ and φ_0 are non-zero.

Under A 4.1, it follows from theorem 1.4 of Böttcher et al. (2010a) that the inner and large eigenvalues of A_m are bounded away from zero, uniformly in m, i. e., given a small $\varepsilon > 0$ there exists $C_\varepsilon > 0$ such that

$$\alpha_{m,l} \geq C_\varepsilon \tag{4.13}$$

whenever $(m-l+1)/(m+2) \geq \varepsilon$. Further, their theorem 1.5 states that for the small eigenvalues it holds that

$$\alpha_{m,l} = \frac{\lambda''(\pi/2)\pi^2}{8}\left(\frac{m-l+1}{m+2}\right)^2 + O\left(\left(\frac{m-l+1}{m+2}\right)^3\right) \tag{4.14}$$

as $m \to \infty$ and $(m-l)/m \to 0$.

Estimates for the matrix product. The eigenvalues $\beta_{m,0} \geq \ldots \geq \beta_{m,m}$ of $\mathrm{C}_m\mathrm{A}_m$ for a general value of $\gamma > -1/2$ cannot be determined explicitly from those of C_m and A_m alone, in general. However, several useful bounds in terms of the eigenvalues $\alpha_{m,0} \geq \ldots \geq \alpha_{m,m}$ of A_m as well as the estimates on the $c_{m,l}$ can be devised. In particular, we have that

$$\frac{\Gamma(2\gamma)}{\Gamma(\gamma)^2 4^{\gamma-1}} \frac{\alpha_{m,m}}{m+1}(1+o(1)) \overset{(\geq)}{\leq} \beta_{m,m} \overset{(\geq)}{\leq} \frac{\Gamma(2\gamma)}{\Gamma(\gamma)} \frac{\alpha_{m,m}}{(m+1)^\gamma}(1+o(1)), \quad -1/2 < \gamma \overset{(\gamma>1)}{\leq} 1, \tag{4.15}$$

which follows from (4.10) and general bounds on the eigenvalues of products of positive definite Hermitian matrices, see for instance Wang and Zhang (1992) and Zhang and Zhang (2006).

4.3. Minimax estimation

4.3.1. Gaussian white noise sequence models

Review of general infinite white noise sequence models. We start by briefly reviewing some general facts about minimax estimation in infinite white noise sequence models from Cavalier and Tsybakov (2002) and Belitser and Levit (1995). Consider observing

$$Y_k = \theta_k + \varepsilon\sigma_k^{-1}\xi_k, \qquad k = 0,1,2,\dots, \tag{4.16}$$

with $(\xi_k)_k$ an i.i.d. Gaussian white noise, $\varepsilon > 0$ the noise level, and $(\sigma_k)_k$ a known sequence of strictly positive weights. The goal is to estimate the parameter $\theta = (\theta_0, \theta_1, \dots)$ from the noisy observations Y_k. Certainly, estimating θ gets more involved the smaller the weights σ_k are. Asymptotics in this infinite sequence model are w.r.t. $\varepsilon \to 0$.

A *linear estimator* $\hat\theta = \hat\theta(h)$ of θ is defined as $\hat\theta_k = h_k Y_k$ for some given real sequence $h = (h_0, h_1, \dots)$, not depending on the Y_k. The class of linear estimators thus corresponds to the class of real, countably infinite sequences h. The *mean squared risk* of an estimator $\hat\theta$ is defined as

$$R_\varepsilon(\hat\theta, \theta) = \mathrm{E}\|\hat\theta - \theta\|^2 = \sum_{k=0}^{\infty} \mathrm{E}\big[(\hat\theta_k - \theta_k)^2\big].$$

Define the *linear minimax risk* on a class Θ by

$$r_\varepsilon^L(\Theta) = \inf_{h \in \mathbb{R}^{\mathbb{N}}} \sup_{\theta \in \Theta} R_\varepsilon(\hat\theta(h), \theta),$$

and the *minimax risk* on Θ by

$$r_\varepsilon(\Theta) = \inf_{\hat\theta} \sup_{\theta \in \Theta} R_\varepsilon(\hat\theta, \theta),$$

where $\inf_{\hat\theta}$ is the infimum over all possible estimators. An estimator $\hat\theta$ is said to be *rate optimal on* Θ if

$$\sup_{\theta\in\Theta} R_\varepsilon(\hat\theta, \theta) \asymp r_\varepsilon(\Theta) \qquad \text{as } \varepsilon \to 0.$$

It is said to be *asymptotically minimax* or *asymptotically efficient on* Θ if

$$\sup_{\theta\in\Theta} R_\varepsilon(\hat\theta, \theta) \sim r_\varepsilon(\Theta) \qquad \text{as } \varepsilon \to 0.$$

The class Θ is typically chosen to be an l_2-ellipsoid, i. e., given a constant $L > 0$ and a sequence $a = (a_0, a_1, \dots)$ of real ellipsoid weights, set

$$\Theta = \Theta(a, L) = \Big\{\theta : \sum_{k=0}^{\infty} a_k^2 \theta_k^2 \le L\Big\}. \tag{4.17}$$

Pinsker estimator. Let $\Theta = \Theta(a, L)$ be an ellipsoid according to (4.17), and assume that for all $\varepsilon > 0$ there exists a solution c_ε to the equation

$$\varepsilon^2 \sum_{k=0}^{\infty} \sigma_k^{-2} a_k (1 - c_\varepsilon a_k)_+ = c_\varepsilon L, \tag{4.18}$$

where the subscript $+$ denotes positive part, $x_+ = \max\{x, 0\}$. Then, the *Pinsker estimator* is defined as the linear estimator $\hat\theta(h^*)$ with weights $h_k^* = (1 - c_\varepsilon a_k)_+$, $k = 0, 1, \dots$.

Theorem 4.2 (Pinsker, 1980)**.** *a. The Pinsker estimator $\hat\theta(h^*)$ is linear minimax on $\Theta(a, L)$, i. e.,*

$\sup_{\theta\in\Theta} R_\varepsilon(\hat\theta(h^*),\theta) = r_\varepsilon^L(\Theta)$ *for all* $\varepsilon > 0$*, where the linear minimax risk is given by*

$$r_\varepsilon^L(\Theta) = \varepsilon^2 \sum_{k=0}^{\infty} \sigma_k^{-2}(1-c_\varepsilon a_k)_+. \tag{4.19}$$

b. If

$$\frac{\max_{k:a_k<T}\sigma_k^{-2}}{\sum_{k:a_k<T}\sigma_k^{-2}} \longrightarrow 0 \tag{4.20}$$

as $T\to\infty$*, then* $r_\varepsilon(\Theta)\sim r_\varepsilon^L(\Theta)$ *as* $\varepsilon\to 0$*, i. e., under (4.20) the Pinsker estimator is even asymptotically efficient on* $\Theta(a,L)$*.*

The condition (4.20) is from Cavalier and Tsybakov (2002). As we shall see below, the Pinsker estimator may also be efficient if this condition is not satisfied.

Remark. If the sequence a is monotonically non-decreasing, then there always exists a solution c_ε to (4.18) so that the Pinsker estimator is well-defined and Theorem 4.2 applies. Even more, in this case c_ε is unique and known to be given by

$$c_\varepsilon = \frac{\sum_{k=0}^{N_\varepsilon}\sigma_k^{-2}a_k}{L/\varepsilon^2+\sum_{k=0}^{N_\varepsilon}\sigma_k^{-2}a_k^2},$$

where

$$N_\varepsilon = \max\{k : a_k \le c_\varepsilon^{-1}\} = \max\Big\{n : \varepsilon^2\sum_{k=0}^{n}\sigma_k^{-2}a_k(a_n-a_k)\le L\Big\}, \tag{4.21}$$

and the minimax risk is attained at $(\hat\theta(h^*),\theta^*)$ with

$$\theta_k^* = \frac{\varepsilon}{\sigma_k}\sqrt{\frac{(1-c_\varepsilon a_k)_+}{c_\varepsilon a_k}}, \tag{4.22}$$

see e. g. Belitser and Levit (1995).

The sequence model for the Radon transform. Suppose now that we observe a function $f\in L_2(B_1(0);\mu_2)$ in a white noise model. Evaluating at the singular functions $\Phi_{m,l}$ as given in Theorem 4.1, we obtain the doubly indexed sequence of observations

$$\tilde Y_{m,l} = \langle \mathrm{R}f,\Phi_{m,l}\rangle_{\mu_1} + \varepsilon W(\Phi_{m,l}) = \sigma_{m,l}\,\theta_{m,l} + \varepsilon\xi_{m,l},$$

where $\theta_{m,l} = \langle f,\Psi_{m,l}\rangle_{\mu_2}$ are the Fourier coefficients of f w.r.t. the basis $(\Psi_{m,l})$, and $\xi_{m,l} = W(\Phi_{m,l})$ are independent standard-normal random variables. Now rescale $Y_{m,l} = \sigma_{m,l}^{-1}\tilde Y_{m,l}$, so that

$$Y_{m,l} = \theta_{m,l} + \varepsilon\sigma_{m,l}^{-1}\xi_{m,l}, \qquad m\ge l\ge 0. \tag{4.23}$$

We investigate estimation of θ over the ellipsoids

$$\Theta_1 = \Theta_1(\kappa,L) = \Big\{\theta : \sum_{m\ge l\ge 0}(m+1)^{2\kappa}\theta_{m,l}^2 \le L\Big\},$$

$$\Theta_2 = \Theta_2(\kappa,L) = \Big\{\theta : \sum_{m\ge l\ge 0}(m-l+1)^{2\kappa}(l+1)^{2\kappa}\theta_{m,l}^2 \le L\Big\}.$$

Since $m+1 \leq (m-l+1)(l+1) \leq (m+1)^2$ for any $0 \leq l \leq m$,

$$\Theta_1(2\kappa, L) \subset \Theta_2(\kappa, L) \subset \Theta_1(\kappa, L). \tag{4.24}$$

The ellipsoid Θ_2 was proposed by Johnstone and Silverman (1990) in the context of density estimation. Johnstone (1989) shows that in case of $\gamma = 1$ and $\lambda = 1$ it corresponds to a class of functions having 2κ weak derivatives in a weighted L_2-space, see Proposition 4.1. A simpler yet natural choice is the ellipsoid Θ_1. For further discussion see Section 4.3.4.

In order to apply Pinkser's Theorem 4.2 to these ellipsoids in the doubly-indexed sequence model, we require total orderings $\prec_i$, $i = 1,2$, of the index set $\{(m,l),\ m \geq l \geq 0\}$, for which the weights in Θ_i are non-decreasing: For Θ_1, we let $(m,l) \prec_1 (\tilde{m},\tilde{l})$ if $m < \tilde{m}$ or if $m = \tilde{m}$ and $l < \tilde{l}$. Similarly, for Θ_2 we let $(m,l) \prec_2 (\tilde{m},\tilde{l})$ if $(l+1)(m-l+1) < (\tilde{l}+1)(\tilde{m}-\tilde{l}+1)$ or if there is equality and $l < \tilde{l}$.

4.3.2. Limited angle Radon transform

We start with estimation in the limited angle case, where $\lambda = \mathbf{1}_{[-\eta,\eta]}$ for an $\eta < \pi/2$ and hence where the minimal eigenvalue $\alpha_{m,m}$ of A_m decays exponentially according to (4.11). By (4.15), this implies exponential decay of $\sigma_{m,m}$ as well.

In this case condition (4.20) fails to hold and therefore the second part of Pinsker's Theorem 4.2 as stated above does not apply. For a single indexed white noise model, in which the σ_k decay at precise exponential rates, Belitser and Levit (1995) and Golubev and Khasminskii (1999) show that the Pinsker estimator remains asymptotically minimax over Sobolev-type smoothness classes. However, since no general results are available, we start from scratch and give a specifically tailored result for minimax rates in severely ill-posed, doubly indexed sequence models, where in particular the rate of decay of $\sigma_{m,m}$ is only known up to a polynomial factor.

We define the projection estimator $\hat{\theta}(h^{Pr})$ with truncation level M_ε as the linear estimator with $h_{m,l} = 1$ for all $0 \leq l \leq m \leq M_\varepsilon$, and $h_{m,l} = 0$ otherwise.

Theorem 4.3. *If there exist $\rho_1, \rho_2 \in \mathbb{R}$ and $\tau_1 \geq \tau_2 > 0$ such that the sequence of smallest singular values $\sigma_{m,m}$ satisfies*

$$m^{\rho_1} e^{-\tau_1 m} \lesssim \sigma_{m,m} \lesssim m^{\rho_2} e^{-\tau_2 m} \qquad \text{as } m \to \infty, \tag{4.25}$$

then

$$r_\varepsilon(\Theta_i(\kappa,L)) \log(1/\varepsilon)^{2\kappa}(L^{-1} + o(1)) \in [\tau_2^{2\kappa}, \tau_1^{2\kappa}] \qquad \text{as } \varepsilon \to 0, \quad i = 1,2.$$

If in particular $\tau_1 = \tau_2 = \tau$, then any projection estimator $\hat{\theta}(h^{Pr})$ with truncation level

$$M_\varepsilon = \left\lfloor \tau^{-1} \log(1/\varepsilon)(1 - \log(1/\varepsilon)^{-\delta}) \right\rfloor$$

for some $\delta \in (0,1)$ is efficient on $\Theta_i(\kappa,L)$, $i = 1,2$, and the corresponding minimax risk is given by

$$r_\varepsilon(\Theta_i(\kappa,L)) \sim \tau^{2\kappa} L \log(1/\varepsilon)^{-2\kappa} \qquad \text{as } \varepsilon \to 0.$$

The latter result now provides the minimax rate for the limited angle tomography problem for any $\gamma > -1/2$. Indeed, in view of (4.11) as well as the bound given in (4.15),

$$m^{-1/4} e^{-\xi m/2} \overset{(\gtrsim)}{\lesssim} \sigma_{m,m} \overset{(\gtrsim)}{\lesssim} e^{-\xi m/2} m^{1/4-\gamma/2}, \qquad -1 \overset{(\gamma>1)}{<} \gamma \leq 1,$$

and we readily arrive at

Corollary 4.1. *For any $\gamma > -1/2$, the limited angle tomography problem with $\eta < \pi/2$ has minimax risk*

$$r_\varepsilon(\Theta_i(\kappa, L)) \sim (\xi/2)^{2\kappa} L \log(1/\varepsilon)^{-2\kappa} \qquad \text{as } \varepsilon \to 0, \quad i = 1, 2,$$

where ξ is given in (4.12).

Remark. 1. The projection estimator is asymptotically efficient and does not depend on the parameters κ and L of the smoothness class Θ_i, it is thus adaptive. Since the projection estimator is linear and the Pinsker estimator linear minimax (for fixed ε), the Pinsker estimator is of course also efficient.

2. Belitser and Levit (1995) consider a single indexed sequence model where $\sigma_k^{-2} = e^{\beta k^r}$ with $\beta, r > 0$, see their example 6. They show that the Pinsker estimator is efficient. However, it depends on the smoothness index, though.

3. Golubev and Khasminskii (1999) also investigate a single indexed sequence model, in which $\sigma_k^{-2} = e^{\alpha k}/k$ for an $\alpha > 0$. They show that the Pinsker estimator is even second order minimax, the second order term being of order $\sim \log\log \varepsilon^{-2}/(\log \varepsilon^{-2})^{2\kappa+1}$, where the parameter κ corresponds to the smoothness class. Analogous results in our model appear to be difficult, since the singular values are less precisely known.

Finally, we show that the logarithmic rate remains true for general λ (not necessarily an indicator function) which vanishes on an interval at the boundaries.

Corollary 4.2. *Let the weight function $\lambda : [-\pi/2, \pi/2] \to [0, \infty)$ be Lebesgue measurable and bounded above. If there exist $0 < \eta_1 < \eta_2 < \pi/2$ such that*

$$\inf_{|\varphi| \leq \eta_1} \lambda(\varphi) > 0, \qquad \sup_{|\varphi| > \eta_2} \lambda(\varphi) = 0,$$

then

$$r_\varepsilon(\Theta_i(\kappa, L)) \log(1/\varepsilon)^{2\kappa} (2^{2\kappa} L^{-1} + o(1)) \in [\xi_2^{2\kappa}, \xi_1^{2\kappa}] \qquad \text{as } \varepsilon \to 0, \quad i = 1, 2,$$

for any $\gamma > -1/2$, where the ξ_j correspond to η_j according to (4.12).

4.3.3. Bounds on the minimax risk for polynomial decay

In this section we give bounds on the minimax rates in case where the minimal eigenvalue $\alpha_{m,m}$ of the Toeplitz matrix A_m and hence the minimal singular value $\sigma_{m,m}$ decays at a polynomial rate. From the discussion in Section 4.2.1, this is generally the case if λ only has a single root (mod π). We have the following result.

Proposition 4.2. *a. If there exists $\rho \geq 0$ such that $\beta_{m,m} \gtrsim m^{-\rho}$ as $m \to \infty$, then*

$$r_\varepsilon(\Theta_i(\kappa, L)) = O\big(\varepsilon^{\frac{4\kappa}{2\kappa+\rho+2}}\big) \qquad \text{as } \varepsilon \to 0, \quad i = 1, 2.$$

b. Let $C > 0$ and $0 \leq \rho_1 \leq \rho < \rho_1 + 1$. If

$$m^{-\rho} \lesssim \beta_{m,m} \lesssim m^{-\rho_1} \qquad \text{as } m \to \infty, \tag{4.26}$$

then the Pinsker estimator on $\Theta_i(\alpha, L)$ is asymptotically efficient, and

$$r_\varepsilon(\Theta_i(\alpha, L)) \gtrsim \varepsilon^{\frac{4\kappa+2(\rho-\rho_1)}{2\kappa+\rho+1}} \qquad \text{as } \varepsilon \to 0, \quad i = 1, 2.$$

c. If

$$\beta_{m,m}^{-1} \sim C m^{\rho} \qquad \text{as } m \to \infty, \tag{4.27}$$

then

$$r_\varepsilon(\Theta_i(\kappa, L)) \geq \tilde{C} \varepsilon^{\frac{4\kappa}{2\kappa+\rho+1}} (1 + o(1)) \qquad \text{as } \varepsilon \to 0, \quad i = 1, 2,$$

where

$$\tilde{C} = \tilde{C}(\kappa, \rho, L, C) = \Big(\frac{C\kappa}{\pi(\kappa+\rho+1)}\Big)^{\frac{2\kappa}{2\kappa+\rho+1}} \frac{(L(2\kappa+\rho+1))^{\frac{\rho+1}{2\kappa+\rho+1}}}{\rho+1}.$$

Remark. 1. If the minimal eigenvalue $\alpha_{m,m} \asymp m^{-\tilde{\rho}}$, then from the estimate in (4.15), the condition of a. is satisfied with $\rho = \tilde{\rho} + 1$ in case $-1/2 < \gamma \leq 1$, as well as $\rho = \tilde{\rho} + \gamma$ for $\gamma > 1$. Further, (4.26) is satisfied if $0 < \gamma < 2$, in which case ρ is as before and $\rho_1 = \tilde{\rho} + \gamma$ for $0 < \gamma < 1$, and $\rho_1 = \tilde{\rho} + 1$ otherwise. Finally, for condition (4.27) we require $\gamma = 1$.

2. In the next section we shall see that under (4.27) and some additional conditions, the lower bound from c. will actually often be the minimax rate, and further the minimax constant will also be of the form $\tilde{C}(\kappa, \rho, L, C)$ but for distinct values of C, see below. Thus, there are only few minimal eigenvalues in A_m which drive the overall rate of convergence.

4.3.4. Exact minimax rates and efficiency constants in case $\gamma = 1$

Next we intend to find minimax rates and efficiency constants in case where the minimal eigenvalue $\alpha_{m,m}$ and hence the minimal singular value $\sigma_{m,m}$ decays at a polynomial rate. We shall require quite precise asymptotics of all singular values $\sigma_{m,l}$, for which, however, in general only bounds are available, see Section 4.2.1.

Therefore, in this section we restrict ourselves to the case $\gamma = 1$ (fan beam design), so that $\sigma_{m,l} = \sqrt{\pi \alpha_{m,l}/(m+1)}$ as given in (4.8). We shall impose the following assumptions on the eigenvalues $\alpha_{m,l}$ of the Toeplitz matrices A_m, which we show below are satisfied in the banded case.

Assumption A 4.2. There exist $C > 0$ and $\rho \geq 1$ such that

$$\sum_{l=0}^{m} \alpha_{m,l}^{-1} \sim C m^{\rho-1} \qquad \text{as } m \to \infty.$$

Assumption A 4.3. There exist $\rho \geq 2$, $\delta > 0$, and a positive, bounded sequence $c = (c_0, c_1, \ldots)$ such that

$$\alpha_{m,l}^{-1} = c_{m-l} l^{\rho-1} + O\big(((m-l+1)(l+1))^{\rho-1-\delta}\big), \qquad m \geq l \geq 0.$$

Remark. We use the exponent $\rho - 1$ instead of ρ since the parameter ρ then compares to that of Section 4.3.3.

Proposition 4.3. *Suppose that λ is banded and that Assumption A 4.1 holds true. Then the eigenvalues $\alpha_{m,l}$ satisfy Assumption A 4.2 with $\rho = 3$ and $C = 4/(3\lambda''(\pi/2))$, as well as Assumption A 4.3 with $\rho = 3$ and $c_j = \frac{8}{\lambda''(\pi/2)\pi^2}(j+1)^{-2}$.*

Before turning to minimax estimation, we further discuss the smoothness classes in case $\gamma = 1$ based on Proposition 4.1. Introduce

$$\Theta_3 = \Theta_3(\kappa, L) = \Big\{\theta : \sum_{m \geq l, k \geq 0} (m-k+1)^{2\kappa}(k+1)^{2\kappa}\big(v_{m,l}^{(k)}\big)^2 \theta_{m,l}^2 \leq L\Big\},$$

where $\mathrm{v}_{m,l} = (v_{m,l}^{(0)}, \ldots, v_{m,l}^{(m)})'$ are the orthonormal eigenvectors of A_m. By Proposition 4.1, Θ_3 corresponds to functions having 2κ weak derivatives which are bounded by a constant depending on L, in a weighted L_2-space. However, an analytic treatment of Θ_3 is difficult since the behavior of the entries $v_{m,l}^{(k)}$ of the eigenvectors of A_m is generally unknown, and even in the specific cases where results are available (cf. Böttcher, Grudsky and Maksimenko, 2010b), these are pretty involved. We shall therefore focus on the smoothness classes Θ_1 and Θ_2, and only point out the inclusion relations

$$\Theta_1(2\kappa, L) \subset \Theta_3(\kappa, L) \subset \Theta_1(\kappa, L),$$

which follow since $(m+1)^{2\kappa} \leq \sum_{l,k=0}^{m}(m-k+1)^{2\kappa}(k+1)^{2\kappa}\big(v_{m,l}^{(k)}\big)^2 \leq (m+1)^{4\kappa}$ for any $0 \leq l \leq m$.

Linear Minimax risk on Θ_1 under A 4.2. Let $a_{m,l} = (m+1)^{\kappa}$ be the ellipsoid weights corresponding to $\Theta_1(\kappa, L)$. From (4.21) we have

$$(m,l)_\varepsilon = \max\Big\{(\tilde m, \tilde l) : \varepsilon^2 \sum_{(m,l) \prec_1 (\tilde m, \tilde l)} \sigma_{m,l}^{-2} a_{m,l}(a_{\tilde m, \tilde l} - a_{m,l}) \leq L\Big\},$$

where the maximum is taken w.r.t. the total ordering $\prec_1$ defined at the end of Section 4.3.1. Since $a_{m,0} = \ldots = a_{m,m}$ for all m, we may include all l for the maximal value of m (since these do not increase the sum). Therefore, $(m,l)_\varepsilon = (N_\varepsilon, N_\varepsilon)$, where

$$N_\varepsilon = \max\Big\{n : \varepsilon^2 \sum_{m=0}^{n}\sum_{l=0}^{m} \sigma_{m,l}^{-2} a_{m,l}(a_{n,n} - a_{m,l}) \leq L\Big\}.$$

By A 4.2 we have $\sum_{l=0}^{m} \sigma_{m,l}^{-2} \sim C\pi^{-1} m^{\rho}$, yielding

$$\begin{aligned}\sum_{m=0}^{n}\sum_{l=0}^{m} \sigma_{m,l}^{-2} a_{m,l}(a_{n,n} - a_{m,l}) &\sim \frac{C}{\pi}\sum_{m=0}^{n}\big(n^{\kappa} m^{\kappa+\rho} - m^{2\kappa+\rho}\big) \\ &\sim \frac{C}{\pi}\frac{\kappa}{(\kappa+\rho+1)(2\kappa+\rho+1)} n^{2\kappa+\rho+1}\end{aligned}$$

as $n \to \infty$, and thus

$$N_\varepsilon \sim \Big(\frac{\pi L(\kappa+\rho+1)(2\kappa+\rho+1)}{C\kappa\varepsilon^2}\Big)^{1/(2\kappa+\rho+1)} \qquad \text{as } \varepsilon \to 0.$$

Since $c_\varepsilon \sim N_\varepsilon^{-\kappa}$ by (4.21), and minding that $(1-c_\varepsilon a_{m,l})_+ = 0$ for $m > N_\varepsilon$, from Pinsker's theorem we obtain

$$\begin{aligned} r_\varepsilon^L(\Theta_1(\kappa,L)) &\sim \varepsilon^2 \sum_{m=0}^{N_\varepsilon}\sum_{l=0}^{m} \sigma_{m,l}^{-2}(1-N_\varepsilon^{-\kappa}(m+1)^\kappa) \\ &\sim \frac{C\varepsilon^2}{\pi}\sum_{m=0}^{N_\varepsilon}\big(m^\rho - N_\varepsilon^{-\kappa}m^{\kappa+\rho}\big) \\ &\sim \frac{C\varepsilon^2}{\pi}\frac{\kappa}{(\rho+1)(\kappa+\rho+1)}N_\varepsilon^{\rho+1} \\ &\sim C_1^*\,\varepsilon^{\frac{4\kappa}{2\kappa+\rho+1}} \end{aligned} \tag{4.28}$$

with $C_1^* = C_1^*(\kappa,\rho,L,C)$ given in Theorem 4.4 below.

Linear Minimax risk on Θ_2 under A 4.3. In order to simplify calculations, note that the ellipsoid Θ_2 can be rewritten as

$$\Theta_2(\kappa,L) = \Big\{\theta : \sum_{j,k\ge 0}(j+1)^{2\kappa}(k+1)^{2\kappa}\theta_{j+k,k}^2 \le L\Big\},$$

corresponding to the sequence of ellipsoid weights $a_{j+k,k} = (j+1)^\kappa(k+1)^\kappa$, $j,k \ge 0$. Assumption A 4.3 then reads

$$\alpha_{j+k,k}^{-1} = c_j k^{\rho-1} + O\big(((j+1)(k+1))^{\rho-1-\delta}\big), \qquad j,k \ge 0. \tag{4.29}$$

Define the totally ordered index sets

$$(n) = \big\{(j,k) \in \mathbb{N}_0^2 : (j+1)(k+1) \le n\big\}, \qquad n \in \mathbb{N}.$$

Similarly as above, for the parameter $(j,k)_\varepsilon$ in (4.21) we have $\big\{(j,k) \prec_2 (j,k)_\varepsilon\big\} \cup \{(j,k)_\varepsilon\} = (N_\varepsilon)$, where

$$N_\varepsilon = \max\Big\{n : \varepsilon^2 \sum_{(j,k)\in(n)} \sigma_{j+k,k}^{-2}a_{j+k,k}\big(n^\kappa - a_{j+k,k}\big) \le L\Big\}.$$

Since $\sigma_{j+k,k}^{-2} = (j+k+1)\pi^{-1}\alpha_{j+k,k}^{-1}$, Lemma 4.6 in Section 4.5.3 gives

$$\sum_{(j,k)\in(n)} \sigma_{j+k,k}^{-2}a_{j+k,k}\big(n^\kappa - a_{j+k,k}\big) \sim \frac{K(\rho,c)}{\pi}\frac{\kappa}{(\kappa+\rho+1)(2\kappa+\rho+1)}n^{2\kappa+\rho+1}$$

as $n \to \infty$, where

$$K(\rho,c) = \sum_{j=0}^{\infty} c_j (j+1)^{-(\rho+1)}. \tag{4.30}$$

Therefore,

$$N_\varepsilon \sim \Big(\frac{\pi L(\kappa+\rho+1)(2\kappa+\rho+1)}{K(\rho,c)\kappa\varepsilon^2}\Big)^{1/(2\kappa+\rho+1)} \qquad \text{as } \varepsilon \to 0,$$

so following the lines in (4.28) and using Lemma 4.6, we find that

$$\begin{aligned} r_\varepsilon^L(\Theta_2(\kappa,L)) &= \varepsilon^2 \sum_{(j,k)\in(N_\varepsilon)} \sigma_{j+k,k}^{-2}(1-N_\varepsilon^{-\kappa}a_{j+k,k}) \\ &\sim C_2^*\,\varepsilon^{\frac{4\kappa}{2\kappa+\rho+1}} \end{aligned} \tag{4.31}$$

with $C_2^* = C_2^*(\kappa,\rho,L,c)$ given in Theorem 4.4 below.

Asymptotic efficiency on Θ_1 *and* Θ_2. Given (4.28) and (4.31), the linear minimax risk on Θ_1 and Θ_2, respectively, we now easily arrive at

Theorem 4.4. *For* $i = 1,2$, *under A 4.2 and A 4.3, respectively,*

$$r_\varepsilon(\Theta_i(\kappa,L)) \sim C_i^*\,\varepsilon^{\frac{4\kappa}{2\kappa+\rho+1}} \qquad \textit{as } \varepsilon \to 0,$$

where

$$\begin{aligned} C_i^* &= \Big(\frac{\Xi_i\kappa}{\pi(\kappa+\rho+1)}\Big)^{\frac{2\kappa}{2\kappa+\rho+1}} \frac{\big(L(2\kappa+\rho+1)\big)^{\frac{\rho+1}{2\kappa+\rho+1}}}{\rho+1}, \\ \Xi_i &= \begin{cases} C, & i=1, \\ K(\rho,c), & i=2. \end{cases} \end{aligned} \tag{4.32}$$

Example 4.1. For the ordinary Radon transform, i. e. $\lambda = 1$, we have $\sum_{l=0}^{m} \alpha_{m,l}^{-1} = m+1$, whence A 4.2 is satisfied for $C = 1$ and $\rho = 2$, leading to the minimax rate

$$r_\varepsilon(\Theta_1(\kappa,L)) \asymp \varepsilon^{\frac{4\kappa}{2\kappa+3}} \qquad \text{as } \varepsilon \to 0.$$

On the other hand, Cavalier and Tsybakov (2002) proved that in this case we have

$$r_\varepsilon(\Theta_2(\kappa,L)) \asymp \varepsilon^{\frac{4\kappa}{2\kappa+2}} \qquad \text{as } \varepsilon \to 0,$$

so we apparently improve by estimating within the smaller ellipsoid Θ_2. This is no longer true in general, however, when the inverse problem gets more ill-posed. Consider a banded weight function λ satisfying Assumption A 4.1. Proposition 4.3 and Theorem 4.4 imply that

$$r_\varepsilon(\Theta_i(\kappa,L)) \asymp \varepsilon^{\frac{4\kappa}{2\kappa+4}} \qquad \text{as } \varepsilon \to 0$$

for both $i = 1$ and $i = 2$. A slight improvement can only be found for the efficiency constant. Here, $\Xi_1 = 4/(3\lambda''(\pi/2))$ and $\Xi_2 = 8/(\lambda''(\pi/2)\zeta(6)\pi^2)$, where ζ denotes the Riemann zeta function. Thus, $\Xi_1/\Xi_2 = \pi^2\zeta(6)/6 \approx 1.63$.

4.4. Concluding remarks

1. We have shown how the design influences the degree of ill-posedness of the Radon transform in two dimensions, and that the whole range from mildly ill-posedness to severely

ill-posedness may arise quite naturally.

2. Without weight on the angle, the rate of convergence remains the same over $\Theta_1(\kappa, L)$ for all parameters $\gamma \in (0,1]$ (which governs the weight function on the signed distance), see Section 4.6.5 in the supplement, where we also derive the asymptotic minimax constants.
3. In higher dimensions, injectivity of the limited angle Radon transform as well as the analytic form of its SVD seems not be established.

4.5. Proofs

4.5.1. Proofs of Section 4.3.2

The method of proof for the lower bound resembles that used in Golubev and Khasminskii (1999). Since the proof of proposition 2 in that paper seems to be problematic (in particular the estimate in (26)), we provide a complete proof of a slightly stronger result (see Lemma 4.2 below). The main ingredient is the following lemma.

Lemma 4.1. *Let* $\mu \geq 0, \sigma > 0$, $\mathrm{P}(X = \mu) = \mathrm{P}(X = -\mu) = 1/2$ *and* $Y|X \sim \mathcal{N}(X, \sigma^2)$. *Then*

$$\mathrm{E}\big(\mathrm{E}(X|Y) - X\big)^2 \geq \mu^2\left(1 - 2\mu^2/\sigma^2\right).$$

Proof. We have

$$\mathrm{E}[X|Y] = \mu\,\frac{e^{-\frac{1}{2}\frac{(Y-\mu)^2}{\sigma^2}} - e^{-\frac{1}{2}\frac{(Y+\mu)^2}{\sigma^2}}}{e^{-\frac{1}{2}\frac{(Y-\mu)^2}{\sigma^2}} + e^{-\frac{1}{2}\frac{(Y+\mu)^2}{\sigma^2}}} = \mu\,\frac{e^{\frac{\mu Y}{\sigma^2}} - e^{-\frac{\mu Y}{\sigma^2}}}{e^{\frac{\mu Y}{\sigma^2}} + e^{-\frac{\mu Y}{\sigma^2}}}.$$

Since $\mathrm{E}[X|Y]|(X=\mu) \stackrel{d}{=} -\mathrm{E}[X|Y]|(X=-\mu)$, it follows that

$$\mathrm{E}\big[(\mathrm{E}[X|Y] - X)^2\big] = \mathrm{E}\big[(\mathrm{E}[X|Y] - \mu)^2|X = \mu\big] = \mu^2\,\mathrm{E}\big[4(1+\exp(2Z))^{-2}\big],$$

where $Z \sim \mathcal{N}(t,t)$ with $t = \mu^2/\sigma^2$. It remains to show that

$$\mathrm{E}\big[4(1+\exp(2Z))^{-2}\big] \geq 1 - 2t. \tag{4.33}$$

For any $x \in \mathbb{R}$, $4(1+e^x)^{-2} \geq 3 \cdot \mathbf{1}_{(-\infty,-2]}(x) + (1-x)\cdot \mathbf{1}_{(-2,\infty)}(x)$. Integrating this w.r.t. the distribution of $2Z$ thus gives the lower bound

$$1 + 2\Phi\Big(-\frac{1+t}{\sqrt{t}}\Big) - \int_{-2}^{\infty} \frac{x}{2\sqrt{2\pi t}}\,e^{-\frac{1}{2}\frac{(x-2t)^2}{4t}} = 1 - 2t - R(t)$$

with remainder

$$R(t) = \int_{-\infty}^{-2} \frac{-x}{2\sqrt{2\pi t}}\,e^{-\frac{1}{2}\frac{(x-2t)^2}{4t}} - 2\Phi\Big(-\frac{1+t}{\sqrt{t}}\Big) = \int_{-\infty}^{-2} \frac{-x-2}{2\sqrt{2\pi t}}\,e^{-\frac{1}{2}\frac{(x-2t)^2}{4t}},$$

where Φ is the distribution function of $\mathcal{N}(0,1)$. Evidently, from the last expression it follows that $R(t)$ is non-negative for all $t > 0$, which proves the lower bound (4.33) and thus concludes the proof. □

Lemma 4.2. *For any ellipsoid Θ, the minimax risk in sequence model (4.16) satisfies*

$$r_\varepsilon(\Theta) \geq \sum_k \theta_k^2 - \frac{2}{\varepsilon^2}\sum_k \theta_k^4 \sigma_k^2,$$

uniformly in $\theta = (\theta_k)_{k\geq 0} \in \Theta$ and $\varepsilon > 0$.

Proof. Fix $\theta_0 = (\theta_{0,k})_{k\geq 0} \in \Theta$. Let $\pi_k(\theta_{0,k}) = \pi_k(-\theta_{0,k}) = 1/2$, and let $\pi = \prod_k \pi_k$ be the product distribution on Θ. Then, for all estimators $\hat{\theta}$,

$$\sup_{\theta\in\Theta}\sum_{k=0}^{\infty} \mathrm{E}_\theta\left[(\hat{\theta}_k - \theta_k)^2\right] \geq \int_\Theta \sum_{k=0}^{\infty} \mathrm{E}_\theta\left[(\hat{\theta}_k - \theta_k)^2\right]\pi(d\theta) = \sum_{k=0}^{\infty}\int_\Theta \mathrm{E}_\theta\left[(\hat{\theta}_k - \theta_k)^2\right]\pi(d\theta)$$

and thus

$$r_\varepsilon(\Theta) \geq \sum_{k=0}^{\infty} \inf_{\hat{\theta}_k}\int_\Theta \mathrm{E}_\theta\left[(\hat{\theta}_k - \theta_k)^2\right]\pi(d\theta). \tag{4.34}$$

Now for any $X = (X_k)_{k\geq 0} \sim \pi$ such that $(Y_k, X_k)_{k\geq 0}$ are independent and $Y_k|X_k \sim \mathcal{N}(X_k, \varepsilon^2\sigma_k^{-2})$, by sufficiency, the Bayes risks in (4.34) are minimized by $\hat{\theta}_k = \mathrm{E}[X_k|Y_k]$, so that the conclusion follows from Lemma 4.1. □

Lemma 4.3. *Consider the sequence model (4.16) and the ellipsoid $\Theta(a,L)$ according to (4.17) with $a_k = (k+1)^\kappa$. If there exist $\gamma_1, \gamma_2 > 1$ such that*

$$\liminf_{k\to\infty} \sigma_k/\sigma_{k+1} \geq \gamma_1, \qquad \limsup_{k\to\infty} \sigma_k/\sigma_{k+1} \leq \gamma_2, \tag{4.35}$$

then

$$\varepsilon^{-2}\sum_{k=0}^{\infty}(\theta_k^*)^4\sigma_k^2 = r_\varepsilon^L(\Theta(a,L))O(N_\varepsilon^{-1}) \qquad \text{as } \varepsilon \to 0,$$

where θ^* is the Pinsker solution according to (4.22).

Proof. First, we may rewrite

$$\varepsilon^{-2}\sum_{k=0}^{\infty}(\theta_k^*)^4\sigma_k^2 = \varepsilon^2\sum_{k=0}^{N_\varepsilon}\sigma_k^{-2}\left(\frac{1-c_\varepsilon a_k}{c_\varepsilon a_k}\right)^2,$$

where $c_\varepsilon \sim N_\varepsilon^{-\kappa}$. Set $n_\varepsilon = \lfloor N_\varepsilon/2 \rfloor$, and define the partial sums

$$S_{1,\varepsilon} = \sum_{k=0}^{n_\varepsilon}\sigma_k^{-2}(1-c_\varepsilon a_k)^2/(c_\varepsilon a_k)^2, \qquad S_{2,\varepsilon} = \sum_{k=n_\varepsilon+1}^{N_\varepsilon}\sigma_k^{-2}(1-c_\varepsilon a_k)^2/(c_\varepsilon a_k)^2.$$

The first sum $S_{1,\varepsilon}$ is comparatively small since it comprises the larger σ_k only. In fact, with (4.35) it follows that

$$S_{1,\varepsilon} \leq c_\varepsilon^{-2}\sum_{k=0}^{n_\varepsilon}\sigma_k^{-2} \lesssim \sigma_{N_\varepsilon}^{-2}c_\varepsilon^{-2}\sum_{k=0}^{n_\varepsilon}\gamma_1^{-2(N_\varepsilon-k)} \lesssim \sigma_{N_\varepsilon}^{-2}N_\varepsilon^{2\kappa}\gamma_1^{-N_\varepsilon}$$

which is $O(\sigma_{N_\varepsilon}^{-2} N_\varepsilon^{-\delta})$ for any $\delta > 0$. Using $1-(1-x)^\kappa \le \max(1,\kappa)x$, $0 \le x \le 1$, as well as $c_\varepsilon > a_{N_\varepsilon+1}^{-1}$ and (4.35) again, the second sum satisfies

$$\begin{aligned}
S_{2,\varepsilon} &\lesssim \sum_{k=n_\varepsilon+1}^{N_\varepsilon} \sigma_k^{-2}(1-c_\varepsilon a_k)^2 = \sum_{j=0}^{N_\varepsilon-n_\varepsilon-1} \sigma_{N_\varepsilon-j}^{-2}(1-c_\varepsilon a_{N_\varepsilon-j})^2 \\
&\lesssim \sigma_{N_\varepsilon}^{-2} \sum_{j=0}^{N_\varepsilon-n_\varepsilon-1} \gamma_1^{-2j}\Big(1-\Big(\frac{N_\varepsilon-j+1}{N_\varepsilon+2}\Big)^\kappa\Big)^2 \lesssim \sigma_{N_\varepsilon}^{-2} \sum_{j=0}^{N_\varepsilon-n_\varepsilon-1} \gamma_1^{-2j}\Big(\frac{j+1}{N_\varepsilon+2}\Big)^2 \\
&\lesssim \sigma_{N_\varepsilon}^{-2} N_\varepsilon^{-2}. \qquad (4.36)
\end{aligned}$$

With this, both sums $S_{1,\varepsilon}$ and $S_{2,\varepsilon}$ can now be bounded above in terms of the linear minimax risk $r_\varepsilon^L(\Theta)$ as follows. Using $c_\varepsilon \le a_{N_\varepsilon}^{-1}$, $1-(1-x)^\kappa \ge \min(1,\kappa)x$, $0 \le x \le 1$, and the second inequality in (4.35),

$$\begin{aligned}
r_\varepsilon^L(\Theta) &= \varepsilon^2 \sum_{j=0}^{N_\varepsilon} \sigma_{N_\varepsilon-j}^{-2}(1-c_\varepsilon a_{N_\varepsilon-j}) \gtrsim \varepsilon^2 \sigma_{N_\varepsilon}^{-2} \sum_{j=0}^{N_\varepsilon} \gamma_2^{-2j}\Big(1-\Big(\frac{N_\varepsilon+1-j}{N_\varepsilon+1}\Big)^\kappa\Big) \\
&\gtrsim \varepsilon^2 \sigma_{N_\varepsilon}^{-2}(N_\varepsilon+1)^{-1} \sum_{j=0}^{N_\varepsilon} \gamma_2^{-2j} j \\
&\gtrsim \varepsilon^2 \sigma_{N_\varepsilon}^{-2} N_\varepsilon^{-1}. \qquad (4.37)
\end{aligned}$$

This provides

$$\varepsilon^2(S_{1,\varepsilon}+S_{2,\varepsilon}) \lesssim \varepsilon^2 \sigma_{N_\varepsilon}^{-2} N_\varepsilon^{-2} \lesssim r_\varepsilon^L(\Theta) N_\varepsilon^{-1}$$

and thus concludes the proof. □

Proof of Theorem 4.3. First we prove that $\tau_2^{2\kappa} L \log(1/\varepsilon)^{-2\kappa}$ is an asymptotic lower bound on the minimax risk on Θ_i. In a second step we calculate the risk of the specific projection estimator as introduced in the theorem and show that it attains the upper bound.

Consider the subellipsoid

$$\tilde{\Theta} = \tilde{\Theta}(\kappa,L) = \Big\{\theta : \sum_{m=0}^{\infty}(m+1)^{2\kappa}\theta_{m,m}^2 \le L, \theta_{m,l}=0, m \ne l\Big\}, \qquad (4.38)$$

and given an estimator $\hat{\theta}$ define the estimator $\tilde{\theta}$ by

$$\tilde{\theta}_{m,l} = \begin{cases} \hat{\theta}_{m,l}, & m = l, \\ 0, & m \ne l. \end{cases}$$

Then, $R_\varepsilon(\hat{\theta},\theta) \ge R_\varepsilon(\tilde{\theta},\theta)$ for all $\theta \in \tilde{\Theta}$, and since $\tilde{\Theta}(\kappa,L) \subset \Theta_i(\kappa,L)$,

$$\sup_{\theta\in\Theta_i} R_\varepsilon(\hat{\theta},\theta) \ge \sup_{\theta\in\tilde{\Theta}} R_\varepsilon(\hat{\theta},\theta) \ge \sup_{\theta\in\tilde{\Theta}} R_\varepsilon(\tilde{\theta},\theta).$$

As $\hat{\theta}$ was arbitrary, this shows that

$$r_\varepsilon(\Theta_i(\kappa,L)) \ge \inf_{\hat{\theta}:\theta_{m,l}=0, m\ne l} \sup_{\theta\in\tilde{\Theta}} R_\varepsilon(\hat{\theta},\theta),$$

where the right-hand side, by Lemma 4.2 and the assumption, is in turn bounded below by

$$\sum_{m=0}^{\infty} \theta_{m,m}^2 - \frac{2}{\varepsilon^2} \sum_{m=0}^{\infty} \theta_{m,m}^4 \sigma_{m,m}^2 \geq \sum_{m=0}^{\infty} \theta_{m,m}^2 - \frac{C}{\varepsilon^2} \sum_{m=0}^{\infty} \theta_{m,m}^4 m^{2\rho_2} e^{-2\tau_2 m}$$

for some $C > 0$, uniformly in $\theta \in \tilde{\Theta}$.

Now, the term on the right can be bounded by the linear minimax risk $\tilde{r}_\varepsilon^L$ corresponding to a sequence model with $\sigma_{m,m}$ replaced by $\tilde{\sigma}_{m,m} = m^{\rho_2} e^{-\tau_2 m}$, for which condition (4.35) is satisfied. In fact, letting $\tilde{\theta}^*$ be the Pinsker solution according to (4.22) corresponding to this surrogate sequence model, we have

$$\sum_{m=0}^{\infty} (\tilde{\theta}_{m,m}^*)^2 \geq \sum_{m=0}^{\infty} \frac{\varepsilon^2 \, \tilde{\sigma}_{m,m}^2}{\tilde{\sigma}_{m,m}^2 \varepsilon^2 + (\tilde{\theta}_{m,m}^*)^2} (\tilde{\theta}_{m,m}^*)^2 = \tilde{r}_\varepsilon^L(\tilde{\Theta}),$$

and from Lemma 4.3 it follows that

$$\varepsilon^{-2} \sum_{m=0}^{\infty} (\tilde{\theta}_{m,m}^*)^4 \tilde{\sigma}_{m,m}^2 = o(\tilde{r}_\varepsilon^L(\tilde{\Theta})), \tag{4.39}$$

which together provide

$$\inf_{\hat{\theta}:\theta_{m,l}=0, m\neq l} \sup_{\theta \in \tilde{\Theta}} R_\varepsilon(\hat{\theta}, \theta) \geq \tilde{r}_\varepsilon^L(\tilde{\Theta})(1 + o(1)).$$

Hence, for the lower bound it remains to evaluate the surrogate linear minimax risk $\tilde{r}_\varepsilon^L(\tilde{\Theta})$.

Denoting by $\tilde{c}_\varepsilon$ and $\tilde{N}_\varepsilon$ the solutions to (4.18) and (4.21) in the surrogate model with $\tilde{\sigma}_{m,m}$, since $\tilde{c}_\varepsilon(m+1)^\kappa \leq 1$ for $m \leq N_\varepsilon$ we estimate

$$\begin{aligned}
\tilde{r}_\varepsilon^L(\tilde{\Theta}) &= \varepsilon^2 \sum_{m=0}^{\infty} \tilde{\sigma}_{m,m}^{-2} (1 - \tilde{c}_\varepsilon (m+1)^\kappa)_+ \\
&= \tilde{c}_\varepsilon^2 L + \varepsilon^2 \sum_{m=0}^{\tilde{N}_\varepsilon} \tilde{\sigma}_{m,m}^{-2} (1 - \tilde{c}_\varepsilon (m+1)^\kappa)_+^2 \\
&\leq \tilde{c}_\varepsilon^2 L + \varepsilon^{-2} \sum_{m=0}^{\infty} (\tilde{\theta}_{m,m}^*)^4 \tilde{\sigma}_{m,m}^2 = \tilde{c}_\varepsilon^2 L + o(\tilde{r}_\varepsilon^L(\tilde{\Theta}))
\end{aligned}$$

by (4.39), so that

$$\tilde{r}_\varepsilon^L(\tilde{\Theta}) \sim \tilde{c}_\varepsilon^2 L \qquad \text{as } \varepsilon \to 0.$$

Using $\tilde{c}_\varepsilon \sim N_\varepsilon^{-\kappa}$ and $\min(1,\kappa)x \le 1-(1-x)^\kappa \le \max(1,\kappa)x$, $0 \le x \le 1$, we get

$$\begin{aligned}
\tilde{c}_\varepsilon L &= \varepsilon^2 \sum_{m=0}^{\tilde{N}_\varepsilon} \tilde{\sigma}_{m,m}^{-2} (m+1)^\kappa (1-\tilde{c}_\varepsilon (m+1)^\kappa) \\
&\sim \varepsilon^2 \sum_{j=0}^{\tilde{N}_\varepsilon} \tilde{\sigma}_{\tilde{N}_\varepsilon - j, \tilde{N}_\varepsilon - j}^{-2} (\tilde{N}_\varepsilon - j)^\kappa \Big(1 - \Big(1 - \frac{j-1}{\tilde{N}_\varepsilon}\Big)^\kappa\Big) \\
&\asymp \varepsilon^2 e^{2\tau_2 \tilde{N}_\varepsilon} \tilde{N}_\varepsilon^{-1} \sum_{j=0}^{\tilde{N}_\varepsilon} e^{-2\tau_2 j} (\tilde{N}_\varepsilon - j)^{\kappa - 2\rho_2} (j-1) \\
&\asymp \varepsilon^2 e^{2\tau_2 \tilde{N}_\varepsilon} \tilde{N}_\varepsilon^{\kappa - 2\rho_2 - 1},
\end{aligned}$$

where the last sum was approximated using Lemma 4.4 below. Therefore, $\tilde{N}_\varepsilon^{2\kappa - 2\rho_2 - 1} e^{2\tau_2 \tilde{N}_\varepsilon} \asymp \varepsilon^{-2}$, which in turn holds true if and only if

$$\tilde{N}_\varepsilon = \tau_2^{-1} \big(\log(1/\varepsilon) + \frac{2\kappa - 2\rho_2 - 1}{2} \log\log(1/\varepsilon) + O(1) \big),$$

and thus $\tilde{N}_\varepsilon \sim \tau_2^{-1} \log(1/\varepsilon)$. This gives

$$\tilde{c}_\varepsilon \sim \tau_2^\kappa \log(1/\varepsilon)^{-\kappa} \qquad \text{as } \varepsilon \to 0$$

and hence provides the lower bound.

For the upper bound, consider a projection estimator $\hat{\theta}(h^{Pr})$ with trunctation level M_ε. Its risk is given by

$$R_\varepsilon(\hat{\theta}(h^{Pr}), \theta) = \varepsilon^2 \sum_{m=0}^{M_\varepsilon} \sum_{l=0}^{m} \sigma_{m,l}^{-2} + \sum_{m=M_\varepsilon+1}^{\infty} \sum_{l=0}^{m} \theta_{m,l}^2.$$

Now

$$\begin{aligned}
\sup_{\theta \in \Theta_i} \sum_{m=M_\varepsilon+1}^{\infty} \sum_{l=0}^{m} \theta_{m,l}^2 &\le \sup_{\theta \in \Theta_i} M_\varepsilon^{-2\kappa} \sum_{m=M_\varepsilon+1}^{\infty} \sum_{l=0}^{m} (m+1)^{2\kappa} \theta_{m,l}^2 \le L M_\varepsilon^{-2\kappa}, \\
\sum_{m=0}^{n} \sum_{l=0}^{m} \sigma_{m,l}^{-2} &\le \sum_{m=0}^{n} (m+1) \sigma_{m,m}^{-2} \lesssim \sum_{m=0}^{n} m^{1-2\rho_1} e^{2\tau_1 m} \lesssim n^{1-2\rho_1} e^{2\tau_1 n},
\end{aligned} \tag{4.40}$$

where we used Lemma 4.4 below for the last estimate. Therefore, there exists a constant $C > 0$ such that

$$\sup_{\theta \in \Theta_i} R_\varepsilon(\hat{\theta}(h^{Pr}), \theta) \le C \varepsilon^2 M_\varepsilon^{1-2\rho_1} e^{2\tau_1 M_\varepsilon} + M_\varepsilon^{-2\kappa} L.$$

In order to minimize the bound on the right-hand side, M_ε has to be chosen of order $\log(1/\varepsilon)$, and if we specifically take $M_\varepsilon = \lfloor \tau_1^{-1} \log(1/\varepsilon)(1 - \log(1/\varepsilon)^{-\delta}) \rfloor$ for some $\delta \in (0,1)$, then

$$\varepsilon^2 M_\varepsilon^{1-2\rho_1+2\kappa} e^{2\tau_1 M_\varepsilon} \asymp \frac{\log(1/\varepsilon)^{1-2\rho_1+2\kappa}}{e^{2\log(1/\varepsilon)^{1-\delta}}} \longrightarrow 0,$$

yielding

$$\sup_{\theta \in \Theta_i} R_\varepsilon(\hat{\theta}(h^{Pr}), \theta) \le L M_\varepsilon^{-2\kappa} (1 + o(1)) = \tau_1^{2\kappa} L \log(1/\varepsilon)^{-2\kappa} (1 + o(1)).$$

This finally provides the upper bound and thus concludes the proof. □

Lemma 4.4. *For all $\gamma > 1$ and $\delta, c_1, c_2 \in \mathbb{R}$,*

$$\sum_{j=0}^{n} \gamma^{-j}(n-j)^{c_1}(j+\delta)^{c_2} \sim n^{c_1} \sum_{j=0}^{\infty} \gamma^{-j}(j+\delta)^{c_2} \qquad \text{as } n \to \infty.$$

We provide the proof of the lemma in the supplement.

Proof of Corollary 4.2. Let $\alpha_{m,l}$ be the eigenvalues of the Toeplitz matrices A_m generated by λ. By assumption, there exist constants $c, C > 0$ such that $\lambda \geq c\mathbf{1}_{[-\eta_1,\eta_1]}$ and $\lambda \leq C\mathbf{1}_{[-\eta_2,\eta_2]}$. Denoting by $\alpha_{m,l}^{(j)}$ the eigenvalues of the Toeplitz matrices generated by $\mathbf{1}_{[-\eta_j,\eta_j]}$, $j = 1,2$, it follows that

$$c\alpha_{m,l}^{(1)} \leq \alpha_{m,l} \leq C\alpha_{m,l}^{(2)}, \qquad m \geq l \geq 0,$$

see Grenander and Szegö (1958). Therefore,

$$m^{1/2}e^{-\xi_1 m} \lesssim \alpha_{m,m} \lesssim m^{1/2}e^{-\xi_2 m}$$

with ξ_i correspondingly defined as in (4.12). Using the bound given in (4.15) as well as the first part of Theorem 4.3 finishes the proof. □

4.5.2. Proofs of Section 4.3.3

Proof of Proposition 4.2. a. Because of the inclusion relation (4.24), it suffices to consider $i = 1$. As in Theorem 4.3, consider a projection estimator $\hat{\theta}(h^{Pr})$ with truncation level M_ε. Its bias is estimated in (4.40), while the variance term may be bounded by

$$\sum_{m=0}^{M_\varepsilon}\sum_{l=0}^{m} \sigma_{m,l}^{-2} \leq \sum_{m=0}^{M_\varepsilon}(m+1)\sigma_{m,m}^{-2} \lesssim \sum_{m=0}^{M_\varepsilon} m^{\rho+1} \lesssim M_\varepsilon^{\rho+2}, \tag{4.41}$$

yielding

$$\sup_{\theta\in\Theta_i} R_\varepsilon(\hat{\theta}(h^{Pr}), \theta) \lesssim \varepsilon^2 M_\varepsilon^{\rho+2} + M_\varepsilon^{-2\kappa}.$$

The bound on the right is minimized choosing M_ε of order $\varepsilon^{-2/(2\kappa+\rho+2)}$, which provides the upper bound.

b. Since

$$\frac{\sigma_{n,n}^{-2}}{\sum_{m=0}^{n}\sum_{l=0}^{m}\sigma_{m,l}^{-2}} \leq \frac{\sigma_{n,n}^{-2}}{\sum_{m=0}^{n}\sigma_{m,m}^{-2}} = O(n^{\rho-\rho_1-1}) = o(1), \quad n \to \infty, \tag{4.42}$$

condition (4.20) is satisfied, and the Pinsker estimator is efficient.

Let $\varepsilon > 0$, $i \in \{1,2\}$, and $\hat{\theta}$ be an arbitrary estimator for $\theta \in \Theta_i$. From the reduction scheme introduced at the beginning of the proof of Theorem 4.3, we at once obtain the lower bound

$$r_\varepsilon(\Theta_i) \geq r_\varepsilon(\tilde{\Theta})$$

with reduced ellipsoid $\tilde{\Theta} = \tilde{\Theta}(\kappa, L)$ defined in (4.38).

We can now use Pinsker's theorem to estimate the minimax risk on $\tilde{\Theta}(\kappa, L)$ which evidently coincides with the minimax risk for estimating the single-indexed sequence $(\theta_{0,0}, \theta_{1,1}, \dots)$ within the ellipsoid $\Theta(a, L)$ defined in (4.17) for $a_m = (m+1)^\kappa$. The linear minimax risk on $\tilde{\Theta}$ is therefore given by

$$r_\varepsilon^L(\tilde{\Theta}) = \varepsilon^2 \sum_{m=0}^{N_\varepsilon} \sigma_{m,m}^{-2}(1 - c_\varepsilon (m+1)^\kappa),$$

where

$$N_\varepsilon = \max\left\{n : \varepsilon^2 \sum_{m=0}^{n} \sigma_{m,m}^{-2}(m+1)^\kappa((n+1)^\kappa - (m+1)^\kappa) \le L\right\}$$

and $c_\varepsilon \sim N_\varepsilon^{-\kappa}$. Using $\sum_{m=0}^n m^z \sim (z+1)^{-1} n^{z+1}$ as $n \to \infty$ for all $z \ge 0$,

$$\sum_{m=0}^{n} m^\rho (m+1)^\kappa((n+1)^\kappa - (m+1)^\kappa) \sim \frac{\kappa(n+1)^{2\kappa+\rho+1}}{(\kappa+\rho+1)(2\kappa+\rho+1)}.$$

As $\varepsilon \to 0$, under (4.26) this provides $N_\varepsilon \gtrsim \varepsilon^{-\frac{2}{2\kappa+\rho+1}}$, so that

$$r_\varepsilon^L(\tilde{\Theta}) \gtrsim \varepsilon^2 \sum_{m=0}^{N_\varepsilon} m^{\rho_1}(1 - N_\varepsilon^{-\kappa}(m+1)^\kappa) \gtrsim \varepsilon^2 N_\varepsilon^{\rho_1+1} \gtrsim \varepsilon^{\frac{4\kappa+2(\rho-\rho_1)}{2\kappa+\rho+1}}.$$

Finally, (4.42) shows that condition (4.20) is satisfied for the sub-problem with $\tilde{\Theta}(\alpha, L)$ as well, so that

$$r_\varepsilon(\tilde{\Theta}(\alpha, L)) \sim r_\varepsilon^L(\tilde{\Theta}(\alpha, L)).$$

c. Under (4.27) we find the exact rates

$$N_\varepsilon \sim \left(\frac{\pi L(\kappa+\rho+1)(2\kappa+\rho+1)}{C\kappa\varepsilon^2}\right)^{\frac{1}{2\kappa+\rho+1}}$$

and

$$\begin{aligned} r_\varepsilon^L(\tilde{\Theta}) &\sim \frac{C\varepsilon^2}{\pi} \sum_{m=0}^{N_\varepsilon} m^\rho (1 - N_\varepsilon^{-\kappa}(m+1)^\kappa) \sim \frac{C\kappa\varepsilon^2 N_\varepsilon^{\rho+1}}{\pi(\rho+1)(\kappa+\rho+1)} \\ &\sim \tilde{C}(\kappa, \rho, L, C)\, \varepsilon^{\frac{4\kappa}{2\kappa+\rho+1}}. \end{aligned}$$

□

4.5.3. Proofs of Section 4.3.4

Lemma 4.5. *If λ is banded and Assumption A 4.1 holds, then there exists a constant $C > 0$ such that the eigenvalues $\alpha_{m,l}$ of the Toeplitz matrices A_m generated by λ satisfy*

$$\left|\alpha_{m,l}^{-1} - \frac{8}{\lambda''(\pi/2)\pi^2}\left(\frac{m+2}{m-l+1}\right)^2\right| \le C\frac{m+2}{m-l+1}, \qquad m \ge l \ge 0.$$

Proof. Set $c = 8/(\lambda''(\pi/2)\pi^2)$ and $\Delta_{m,l} = \left|\alpha_{m,l}^{-1} - c\left(\frac{m+2}{m-l+1}\right)^2\right|$. For the small eigenvalues $\alpha_{m,l}$, (4.14) provides

$$\begin{aligned}\Delta_{m,l} &= \frac{(m-l+1)^2 - c\alpha_{m,l}(m+2)^2}{\alpha_{m,l}(m-l+1)^2} \\ &= \frac{(m-l+1)^2 O\big((m-l+1)/(m+2)\big)}{(m-l+1)^4/(m+2)^2\big(1+O\big((m-l+1)/(m+2)\big)\big)} \\ &= \frac{m+2}{m-l+1}\frac{O(1)}{1+O\big((m-l+1)/(m+2)\big)}.\end{aligned}$$

Choosing $\varepsilon > 0$ small enough, $1+O\big((m-l+1)/(m+2)\big)$ is bounded away from 0, uniformly in m and l, whenever $(m-l+1)/(m+2) \leq \varepsilon$, which shows that there is $C_1 > 0$ such that

$$\Delta_{m,l} \leq C_1(m+2)/(m-l+1), \qquad (m-l+1)/(m+2) \leq \varepsilon.$$

Choosing C_ε according to (4.13), for the inner and large eigenvalues we even obtain the uniform bound

$$\Delta_{m,l} \leq C_\varepsilon^{-1} + c\varepsilon^{-2} =: C_2, \qquad (m-l+1)/(m+2) \geq \varepsilon.$$

Setting $C = \max\{C_1, C_2\}$ concludes the proof. □

Remark. In order to obtain (4.13) and (4.14) we actually apply theorems 1.4 and 1.5 of Böttcher et al. (2010a) to the generating function $g(\varphi) = \lambda(\varphi/2 - \pi/2)$. Due to the additional shift of $\pi/2$, the resulting Toeplitz matrix does not coincide with A_m, it does have the same eigenvalues, though.

Proof of Proposition 4.3. In order to show the statement concerning Assumption A 4.2, in view of Lemma 4.5,

$$\sum_{l=0}^{m} \alpha_{m,l}^{-1} = \frac{8(m+2)^2}{\lambda''(\pi/2)\pi^2}\sum_{l=0}^{m}(m-l+1)^{-2} + \sum_{l=0}^{m} O\Big(\frac{m+2}{m-l+1}\Big).$$

The error is $O(m\log m) = o(m^2)$. Using that $\sum_{j=1}^{\infty} j^{-2} = \pi^2/6$, the driving part is asymptotically equivalent to $\frac{4}{3}m^2/\lambda''(\pi/2)$, concluding the proof.

Concerning Assumption A 4.3, from Lemma 4.5 there exists $C > 0$ such that, for all $m \geq l \geq 0$,

$$\big|\alpha_{m,l}^{-1} - c_{m-l}l^2\big| \leq C(m+2) + \frac{8}{\lambda''(\pi/2)\pi^2}\frac{\big|(m+2)^2 - l^2\big|}{(m-l+1)^2}.$$

Now, $(m+2)^2 = (m-l+1)^2 + 2(m-l+1)(l+1) + l^2 + 2l + 2$, which shows that the right summand is bounded by $C_1(l+1)$ for an adequate constant $C_1 > 0$. Therefore we obtain

$$\big|\alpha_{m,l}^{-1} - c_{m-l}l^2\big| \leq C(m+2) + C_1(l+1) \leq (C+C_1)(m-l+1)(l+1),$$

whence A 4.3 holds true for any $\delta \leq 1$. □

Lemma 4.6. *If there exist $\beta \geq 1$, $\delta > 0$, and a positive, bounded sequence $c = (c_0, c_1, \dots)$ such that*

$$\alpha_{j+k,k}^{-1} = c_j k^\beta + O\big(((j+1)(k+1))^{\beta-\delta}\big), \qquad j,k \geq 0,$$

then, for all $\alpha \geq 0$,

$$\sum_{(j,k)\in(n)} (j+k+1)(j+1)^{\alpha}(k+1)^{\alpha}\alpha_{j+k,k}^{-1} \sim \frac{K(\beta+1,c)}{\alpha+\beta+2}\, n^{\alpha+\beta+2}$$

as $n \to \infty$, *where* $K(\beta,c) = \sum_{j=0}^{\infty} c_j (j+1)^{-(\beta+1)}$.

We provide the proof of the lemma in the supplement.

Proof of Theorem 4.4. In view of (4.28) and (4.31), it remains to show that condition (4.20) holds.

Under A 4.2,

$$\sum_{m=0}^{n}\sum_{l=0}^{m}\sigma_{m,l}^{-2} = \frac{1}{\pi}\sum_{m=0}^{n}(m+1)\sum_{l=0}^{m}\alpha_{m,l}^{-1} \asymp n^{\rho+1}$$

and

$$\max_{m=0,\ldots,n}\max_{l=0,\ldots,m}\sigma_{m,l}^{-2} \leq \max_{m=0,\ldots,n}\sum_{l=0}^{m}\sigma_{m,l}^{-2} \asymp n^{\rho}.$$

And under A 4.3,

$$\max_{(j,k)\in(n)}\sigma_{j+k,k}^{-2} = \max_{(j,k)\in(n)}\Big(\frac{j+k+1}{\pi}c_j k^{\rho-1}\Big) + O(n^{\rho-\delta}) = O(n^{\rho}),$$

while Lemma 4.6 shows that

$$\sum_{(j,k)\in(n)}\sigma_{j+k,k}^{-2} \asymp n^{\rho+1}.$$

So, evidently, in both cases (4.20) holds. □

Acknowledgements

The authors would like to thank Thorsten Hohage for helpful discussions.

4.6. Appendix: Technical supplement

4.6.1. Details on the derivation of the SVD

Lemma 4.7. *If* λ *is integrable, the Radon transform* R *as a map between the weighted* L_2*-spaces in (4.1) is continuous with norm*

$$\|\mathrm{R}\|^2 = \sup_{\|f\|_{\mu_2}=1}\|\mathrm{R}f\|_{\mu_1}^2 = \int_{-\pi/2}^{\pi/2}\lambda(\varphi)\,d\varphi.$$

Proof. For $\varphi \in [-\pi/2, \pi/2]$ fixed, define

$$\mathrm{R}_\varphi : L_2(B_1(0);\mu_2) \longrightarrow L_2([-1,1];w_1(s)\,ds) \tag{4.43}$$

by $\mathrm{R}_\varphi f(s) = \mathrm{R}f(\varphi, s)$. This operator has norm $\|\mathrm{R}_\varphi\| = 1$, see Davison (1981, Theorem 1), providing

$$\|\mathrm{R}f\|_{\mu_1}^2 = \int_{-\pi/2}^{\pi/2} \|\mathrm{R}_\varphi f\|_{w_1}^2 \,\lambda(\varphi)\,d\varphi \le \|f\|_{\mu_2}^2 \int_{-\pi/2}^{\pi/2} \lambda(\varphi)\,d\varphi.$$

Further, w_1^{-1} and w_2^{-1} are normalized to one, and $\mathrm{R}_\varphi w_2^{-1} = w_1^{-1}$ for all φ, yielding

$$\|\mathrm{R}\|^2 = \sup_{\|f\|_{\mu_2}=1} \|\mathrm{R}f\|_{\mu_1}^2 = \int_{-\pi/2}^{\pi/2} \lambda(\varphi)\,d\varphi,$$

which was to be shown. □

Lemma 4.8. *The adjoint operator of* R *is given by*

$$\mathrm{R}^* : L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1) \longrightarrow L_2(B_1(0);\mu_2),$$
$$(\mathrm{R}^*g)(x,y) = w_2(x,y)^{-1} \int_{-\pi/2}^{\pi/2} g(\varphi, x\cos\varphi + y\sin\varphi) w_1(x\cos\varphi + y\sin\varphi)\lambda(\varphi)\,d\varphi.$$

Proof. For $\varphi \in [-\pi/2, \pi/2]$ fixed, let the operator R_φ, as in (4.43), be defined by $(\mathrm{R}_\varphi f)(s) = (\mathrm{R}f)(\varphi, s)$. The adjoint R_φ^* of R_φ is then, for $g \in L_2([-1,1];w_1)$, given by

$$(\mathrm{R}_\varphi^* g)(x,y) = w_2(x,y)^{-1} g(x\cos\varphi + y\sin\varphi) w_1(x\cos\varphi + y\sin\varphi),$$

which, applying the rotation $(x,y) = (s\cos\varphi - t\sin\varphi, s\sin\varphi + t\cos\varphi)$, follows from

$$\begin{aligned}
\langle \mathrm{R}_\varphi f, g\rangle_{w_1} &= \int_{-1}^{1} \int_{-\sqrt{1-s^2}}^{\sqrt{1-s^2}} f(s\cos\varphi - t\sin\varphi, s\sin\varphi + t\cos\varphi) g(s)\, w_1(s)\,dt\,ds \\
&= \int_{B_1(0)} f(x,y) g(x\cos\varphi + y\sin\varphi)\, w_1(x\cos\varphi + y\sin\varphi)\,dx\,dy \\
&= \int_{B_1(0)} f(x,y) (\mathrm{R}_\varphi^* g)(x,y)\, w_2(x,y)\,dx\,dy \\
&= \langle f, \mathrm{R}_\varphi^* g\rangle_{w_2}.
\end{aligned}$$

For $g \in L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1)$, defining g_φ on $[-1,1]$ by $g_\varphi(s) = g(\varphi, s)$, this, by definition of R^*, particularly gives

$$(\mathrm{R}^*g)(x,y) = \int_{-\pi/2}^{\pi/2} (\mathrm{R}_\varphi^* g_\varphi)(x,y)\,\lambda(\varphi)\,d\varphi, \tag{4.44}$$

providing

$$\begin{aligned}\langle \mathrm{R}f, g\rangle_{\mu_1} &= \int_{-\pi/2}^{\pi/2} \langle \mathrm{R}_\varphi f, g_\varphi\rangle_{w_1} \lambda(\varphi)\, d\varphi = \int_{-\pi/2}^{\pi/2} \langle f, \mathrm{R}^*_\varphi g_\varphi\rangle_{w_2} \lambda(\varphi)\, d\varphi \\ &= \int_{B_1(0)} f(x,y) \int_{-\pi/2}^{\pi/2} (\mathrm{R}^*_\varphi g_\varphi)(x,y) \lambda(\varphi)\, d\varphi\, w_2(x,y)\, dx\, dy \\ &= \langle f, \mathrm{R}^* g\rangle_{\mu_2},\end{aligned}$$

which shows that R and R^* are adjoint to one another. □

Lemma 4.9. *For $\phi_m = w_1^{-1} C_m^\gamma$ and h integrable on $[-\pi/2, \pi/2]$, the function $g(\varphi, s) = h(\varphi)\phi_m(s)$ satisfies*

$$(\mathrm{RR}^* g)(\varphi, s) = \frac{\phi_m(s)}{C_m^\gamma(1)} \int_{-\pi/2}^{\pi/2} h(\varphi') C_m^\gamma(\cos(\varphi' - \varphi)) \lambda(\varphi')\, d\varphi'.$$

Proof. Using (4.44), for $g \in L_2([-\pi/2, \pi/2] \times [-1,1]; \mu_1)$ we may rewrite

$$(\mathrm{RR}^* g)(\varphi, s) = \int_{-\pi/2}^{\pi/2} (\mathrm{R}_\varphi \mathrm{R}^*_{\varphi'} g_{\varphi'})(s)\, \lambda(\varphi')\, d\varphi'.$$

Now, from theorem 3.1 in Davison and Grunbaum (1981) it follows that

$$(\mathrm{R}_\varphi \mathrm{R}^*_{\varphi'} \phi_m)(s) = \frac{C_m^\gamma(\cos(\varphi' - \varphi))}{C_m^\gamma(1)}\, \phi_m(s), \qquad \varphi, \varphi' \in [-\pi/2, \pi/2],$$

and, by linearity of R_φ and $\mathrm{R}^*_{\varphi'}$, for $g = h\phi_m$ we have

$$(\mathrm{R}_\varphi \mathrm{R}^*_{\varphi'} g_{\varphi'})(s) = h(\varphi')\, (\mathrm{R}_\varphi \mathrm{R}^*_{\varphi'} \phi_m)(s),$$

which together complete the proof. □

Lemma 4.10. *For $m \geq l \geq 0$, let the operator T_m defined on $L_2([-\pi/2, \pi/2]; \lambda(\varphi)d\varphi)$ be given by*

$$(T_m h)(\varphi) = C_m^\gamma(1)^{-1} \int_{-\pi/2}^{\pi/2} h(\varphi') C_m^\gamma(\cos(\varphi' - \varphi)) \lambda(\varphi')\, d\varphi',$$

and let $h_{m,l}(\varphi) = e^{-i(m-2l)\varphi}$. Set $\mathrm{h}_m = (h_{m,0}, \ldots, h_{m,m})'$ and $T_m \mathrm{h}_m = (T_m h_{m,0}, \ldots, T_m h_{m,m})'$. Let C_m and A_m be defined as in (4.3) and (4.4), $\{\mathrm{v}_{m,l} = (v_{m,l}^{(0)}, \ldots, v_{m,l}^{(m)})'\}_{l=0}^m$ be an orthonormal system of eigenvectors of A_m corresponding to eigenvalues $\alpha_{m,l}$, and define matrices $\mathrm{V}_m = (\mathrm{v}_{m,0}, \ldots, \mathrm{v}_{m,m})$ and $\Lambda_m = \mathrm{diag}(\alpha_{m,0}, \ldots, \alpha_{m,m})$. Then the following statements hold:

a. T_m vanishes on the orthogonal complement of $\mathrm{lin}\{h_{m,l}\}_{l=0}^m$.

b. $T_m \mathrm{h}_m = \pi(\mathrm{C}_m \mathrm{A}_m)' \mathrm{h}_m$.

c. The functions

$$\tilde{h}_{m,l} = \frac{1}{\sqrt{\pi \alpha_{m,l}}} \mathrm{v}'_{m,l} \mathrm{h}_m = \frac{1}{\sqrt{\pi \alpha_{m,l}}} \sum_{k=0}^m v_{m,l}^{(k)} h_{m,k}, \qquad l = 0, \ldots, m,$$

are an orthonormal basis of $\mathrm{lin}\{h_{m,l}\}_{l=0}^m$.

d. $T_m\tilde{\mathrm{h}}_m = \mathrm{B}'_m\tilde{\mathrm{h}}_m$*, where* $\tilde{\mathrm{h}}_m = (\tilde{h}_{m,0},\dots,\tilde{h}_{m,m})'$, $T_m\tilde{\mathrm{h}}_m = (T_m\tilde{h}_{m,0},\dots,T_m\tilde{h}_{m,m})'$*, and*

$$\mathrm{B}_m = \pi\Lambda^{1/2}\mathrm{V}^*_m\mathrm{C}_m\mathrm{V}_m\Lambda^{1/2}.$$

Proof. In view of (4.9.19) and (4.9.21) in Szegö (1967), the polynomials $C^\gamma_m(\cos\varphi)$ attain the explicit form

$$C^\gamma_m(\cos\varphi) = \sum_{j=0}^m \frac{\Gamma(j+\gamma)\Gamma(m-j+\gamma)}{\Gamma(\gamma)^2 j!(m-l)!}e^{i(m-2j)\varphi},$$

so that, since $C^\gamma_m(1) = \Gamma(m+2\gamma)/(\Gamma(2\gamma)m!)$, setting

$$c_{m,j} = \binom{m}{j}\frac{\Gamma(2\gamma)\Gamma(j+\gamma)\Gamma(m-j+\gamma)}{\Gamma(m+2\gamma)\Gamma(\gamma)^2}$$

we find that

$$T_m h(\varphi) = \sum_{j=0}^m c_{m,j}e^{-i(m-2j)\varphi}\int_{-\pi/2}^{\pi/2} h(\varphi')e^{i(m-2j)\varphi'}\lambda(\varphi')\,d\varphi'.$$

This evidently shows that $T_m h = 0$ for h in the orthogonal complement of $\mathrm{lin}\{h_{m,0},\dots,h_{m,m}\}$ in $L_2([-\pi/2,\pi/2];\lambda(\varphi)d\varphi)$, which is part a, and defining

$$a_z = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} e^{-i2z\varphi'}\lambda(\varphi')\,d\varphi', \qquad z\in\mathbb{Z},$$

we find that

$$T_m h_{m,l} = \pi\sum_{j=0}^m c_{m,j}a_{l-j}h_{m,j},$$

proving part b.

Orthonormality of the functions $\tilde{h}_{m,0}\dots,\tilde{h}_{m,m}$ follows from that of $\mathrm{v}_{m,0},\dots,\mathrm{v}_{m,m}$. In fact, using

$$\langle h_{m,k_1},h_{m,k_2}\rangle_\lambda = \int_{-\pi/2}^{\pi/2} h_{m,k_1}(\varphi)\overline{h_{m,k_2}(\varphi)}\lambda(\varphi)\,d\varphi = \pi a_{k_2-k_1},$$

we have

$$\begin{aligned}\langle\tilde{h}_{m,l_1},\tilde{h}_{m,l_2}\rangle_\lambda &= \frac{1}{\sqrt{\alpha_{m,l_1}\alpha_{m,l_2}}}\sum_{k_1,k_2=0}^m v^{(k_1)}_{m,l_1}\overline{v^{(k_2)}_{m,l_2}}a_{k_2-k_1}\\ &= \frac{1}{\sqrt{\alpha_{m,l_1}\alpha_{m,l_2}}}\overline{\mathrm{v}'_{m,l_2}}\mathrm{A}_m\mathrm{v}_{m,l_1} = \sqrt{\frac{\alpha_{m,l_1}}{\alpha_{m,l_2}}}\overline{\mathrm{v}'_{m,l_2}}\mathrm{v}_{m,l_1}.\end{aligned}$$

This in particular implies that $\tilde{h}_{m,0},\dots,\tilde{h}_{m,m}$ are linearly independent so that, since we have $\tilde{h}_{m,l}\in\mathrm{lin}\{h_{m,l}\}_{l=0}^m$, $l=0,\dots,m$, they are a corresponding basis, too, concluding part c.

Finally, note that $\tilde{\mathrm{h}}_m = \pi^{-1/2}\Lambda^{-1/2}\mathrm{V}'_m\mathrm{h}_m$, $\mathrm{h}_m = \pi^{1/2}\overline{\mathrm{V}}_m\Lambda^{1/2}\tilde{\mathrm{h}}_m$, and $\mathrm{A}_m\mathrm{V}_m = \mathrm{V}_m\Lambda_m$, with part b providing

$$\begin{aligned}T_m\tilde{\mathrm{h}}_m &= \pi^{-1/2}\Lambda_m^{-1/2}\mathrm{V}'_m T_m\mathrm{h}_m = \pi^{1/2}\Lambda_m^{-1/2}\mathrm{V}'_m\mathrm{A}'_m\mathrm{C}_m\mathrm{h}_m\\ &= \pi^{1/2}\Lambda_m^{1/2}\mathrm{V}'_m\mathrm{C}_m\mathrm{h}_m = \pi\Lambda_m^{1/2}\mathrm{V}'_m\mathrm{C}_m\overline{\mathrm{V}}_m\Lambda_m^{1/2}\tilde{\mathrm{h}}_m,\end{aligned}$$

which shows part d. □

Lemma 4.11. *The functions* $\Psi_{m,l} = \sigma_{m,l}^{-1} \mathrm{R}^* \Phi_{m,l}$ *are given by*

$$\Psi_{m,l}(x,y) = \frac{\sigma_{m,l}}{\pi^{3/2}\sqrt{d_m}} \sum_{k_1,k_2=0}^{m} \frac{w_{m,l}^{(k_1)} v_{m,k_1}^{(k_2)}}{c_{m,k_2}\sqrt{\alpha_{m,k_1}}} \widetilde{\Psi}_{m,k_2}(x,y), \tag{4.45}$$

where

$$\widetilde{\Psi}_{m,l}(x,y) = \frac{h_{m,l}(\theta) J_{m,l}(r)}{w_2(x,y)}, \qquad J_{m,l}(r) = \frac{\pi\Gamma(\gamma+m-l)}{(m-l)!\Gamma(\gamma)} r^{m-2l} P_l^{(\gamma-1,m-2l)}(2r^2-1),$$

$(x,y) = re^{i\theta}$ *and* $P_n^{(\alpha,\beta)}$ *are the Jacobi polynomials. Further, the span of* $(\Psi_{m,l})_{0\le l\le m}$ *is dense in* $L_2(B_1(0);\mu_2)$*, and hence they form an orthonormal basis.*

Proof. From Davison (1983, theorem 3.2),

$$\big(\mathrm{R}\widetilde{\Psi}_{m,l}\big)(\varphi,s) = \pi c_{m,l} h_{m,l}(\varphi)\,\phi_m(s). \tag{4.46}$$

Call the functions on the right side of (4.45) $\widehat{\Psi}_{m,l}(x,y)$. The functions $(\widetilde{\Psi}_{m,l})_{0\le l\le m}$ form an orthogonal basis of $L_2(B_1(0);\mu_2)$. By orthonormality of the vectors $v_{m,l}$ and $w_{m,l}$, it follows that the $(\widehat{\Psi}_{m,l})_{0\le l\le m}$ form an orthogonal basis of $L_2(B_1(0);\mu_2)$ as well. Further by (4.46) and their definition, we have that $\big(\mathrm{R}\hat{\Psi}_{m,l}\big) = \sigma_{m,l}\Phi_{m,l}$. Since the $(\Phi_{m,l})_{0\le l\le m}$ are orthonormal in $L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1)$, it follows that R as an operator between the weighted L_2-spaces in (4.1) is injective. By (4.7), for the functions $\Psi_{m,l} = \sigma_{m,l}^{-1}\mathrm{R}^*\Phi_{m,l}$ we also have that $\big(\mathrm{R}\Psi_{m,l}\big) = \sigma_{m,l}\Phi_{m,l}$, so that $\Psi_{m,l} = \widehat{\Psi}_{m,l}$ by injectivity. □

4.6.2. Singular functions in case $\gamma = 1$

We specialize our results for the singular functions in brain space, (4.45), for the case $\gamma = 1$. Here the weights $c_{m,l}$ have the simple form $c_{m,l} = (m+1)^{-1}$ for all m, so, given the eigenvalues $\alpha_{m,l}$ of A_m, it follows that $\beta_{m,l} = \alpha_{m,l}/(m+1)$, and thus the singular values of the operator R are $\sigma_{m,l} = \sqrt{\pi\alpha_{m,l}/(m+1)}$, $m \ge l \ge 0$. Further, $d_m = 1$ for all m, and

$$C_m^1(s) = U_m(s) = \frac{\sin((m+1)\arccos s)}{\sin\arccos s}$$

are the Chebyshev polynomials of the second kind. Therefore, the singular functions $\Phi_{m,l}$ in detector space reduce to

$$\Phi_{m,l}(\varphi,s) = \frac{2}{\pi}\sqrt{\frac{1-s^2}{\pi\alpha_{m,l}}}\, U_m(s) \sum_{k=0}^{m} v_{m,l}^{(k)} e^{-i(m-2k)\varphi}$$

with $\{\mathrm{v}_{m,l} = (v_{m,l}^{(0)},\dots,v_{m,l}^{(m)})'\}_{l=0}^{m}$ the orthonormal system of eigenvectors of A_m.

The functions $\widetilde{\Psi}_{m,l}$ reduce to the Zernike functions $z_{m,l}$ defined by

$$z_{m,l}(x,y) = Z_m^{m-2l}(r)e^{-i(m-2l)\theta},$$

where $m \geq l \geq 0$ and $(x,y) = re^{i\theta} \in B_1(0)$, and where the radial part Z_m^{m-2l} on the unit interval $[0,1]$ is given by

$$Z_m^n(r) = \sum_{k=0}^{(m-n)/2} \frac{(-1)^k (m-k)!}{k!((m+n)/2-k)!((m-n)/2-k)!} r^{m-2k}$$

for $m-n$ even. The singular function $\Psi_{m,l}$ in (4.45) are then expressed as

$$\Psi_{m,l}(x,y) = \frac{\sqrt{m+1}}{\pi} \sum_{k=0}^{m} v_{m,l}^{(k)} z_{m,k}(x,y), \qquad m \geq l \geq 0. \tag{4.47}$$

4.6.3. Proof of Proposition 4.1

In order to deduce the summability condition of Proposition 4.1, similar as in Johnstone (1989) we differentiate the singular functions $\Psi_{m,l}$ given in (4.47) by means of the differential operators $D = (\partial/\partial x - i\partial/\partial y)/2$ and $\bar{D} = (\partial/\partial x + i\partial/\partial y)/2$. These differential operators have the advantage of providing neat formulas for the derivatives of the Zernike functions $z_{m,l}$. In fact, we will see below that for $p,q \in \mathbb{N}$ such that $p+q=s$ we get

$$D^p \bar{D}^q z_{m,l} = \begin{cases} \frac{s!}{\pi} h_{m-s,l-p}^{s+1}, & m-q \geq l \geq p, \\ 0, & \text{else}, \end{cases} \tag{4.48}$$

where

$$h_{m,l}^{\gamma}(x,y) = \int_{-\pi/2}^{\pi/2} C_m^{\gamma}(x\cos\varphi + y\sin\varphi) e^{-i(m-2l)\varphi} d\varphi, \tag{4.49}$$

and where the norm of these derivatives with respect to μ_3 is explicitly given by

$$\|D^p \bar{D}^q z_{m,l}\|_{\mu_3}^2 = \frac{\pi^{1/2}(s+1)(2s+1)!}{2^{2s+1} s! \Gamma(s+3/2)} \frac{(m-l+p)!(l+q)!}{(l-p)!(m-l-q)!(m+1)}. \tag{4.50}$$

Proof of Proposition 4.1. Evidently, it suffices to show that the summability condition

$$\sum_{m=0}^{\infty} \sum_{l,k=0}^{m} \theta_{m,l}^2 \left(v_{m,l}^{(k)}\right)^2 (m-k+1)^s (k+1)^s < \infty$$

is equivalent to $D^p \bar{D}^q f \in L_2(B_1(0);\mu_3)$ for all $p,q \in \mathbb{N}$ such that $p+q=s$. For this, we first give bounds on the L_p-norms of the Zernike functions above. Clearly,

$$\frac{(m-l+p)!}{(m-l-q)!} \leq (m-l+p)^s \leq (m-l+1)^s s^s, \qquad \frac{(l+q)!}{(l-p)!} \leq (l+q)^s \leq (l+1)^s s^s.$$

Further, $m-l-q+1 \geq (m-l+1)(q+1)^{-1}$ and $l-p+1 \geq (l+1)(p+1)^{-1}$ whenever $m-q \geq l \geq p$, yielding

$$\frac{(m-l+p)!}{(m-l-q)!} \geq (m-l-q+1)^s \geq (m-l+1)^s (s+1)^{-s},$$
$$\frac{(l+q)!}{(l-p)!} \geq (l-p+1)^s \geq (l+1)^s (s+1)^{-s}.$$

Therefore, by (4.50) there exist constants $c_s, C_s > 0$, only depending on $s = p+q$, such that

$$c_s \leq \frac{m+1}{(m-l+1)^s(l+1)^s} \|D^p \bar{D}^q z_{m,l}\|^2_{\mu_3^s} \leq C_s$$

for all $m-q \geq l \geq p$.

Now, expanding f as a Fourier series in the singular functions $\Psi_{m,l}$,

$$f = \sum_{m=0}^{\infty}\sum_{l=0}^{m} \theta_{m,l}\Psi_{m,l} = \pi^{-1}\sum_{m=0}^{\infty}\sqrt{m+1}\sum_{l=0}^{m}\theta_{m,l}\sum_{k=0}^{m} v_{m,l}^{(k)} z_{m,k},$$

whence, using the orthogonality of the $z_{m,k}$ which in turn follows from that of the $\Psi^1_{m,k}$, see (4.51) below, the weak derivatives of f with respect to the operators D and $\bar{D}$ satisfy

$$\begin{aligned} \|D^p\bar{D}^q f\|^2_{\mu_3^s} &= \pi^{-2}\sum_{m=s}^{\infty}(m+1)\sum_{l=0}^{m}\theta^2_{m,l}\sum_{k=p}^{m-q}\left(v_{m,l}^{(k)}\right)^2\|D^p\bar{D}^q z_{m,k}\|^2_{\mu_3^s} \\ &\asymp \sum_{m=s}^{\infty}\sum_{l=0}^{m}\theta^2_{m,l}\sum_{k=p}^{m-q}\left(v_{m,l}^{(k)}\right)^2(m-k+1)^s(k+1)^s. \end{aligned}$$

This sum is finite for all $p, q \in \mathbb{N}$ such that $p+q = s$ if and only if the same holds true for k ranging from 0 to m. And finally, since the $\theta^2_{m,l}$ are finite due to $f \in L_2$, the outer sum can be extended to m ranging from 0 to infinity. □

Proof of (4.50). For clarity, in the following we express the dependence of all functions on the parameter γ. Further, recall that the measures μ_i^γ, $i = 1,2,3$, are defined in terms of the weight functions

$$\begin{aligned} w_1^\gamma(\varphi,s) &= \frac{\pi^{1/2}\Gamma(\gamma+1/2)}{\gamma\Gamma(\gamma)}(1-s^2)^{1/2-\gamma}, \qquad |s| \leq 1,\ |\varphi| \leq \pi/2, \\ w_2^\gamma(x,y) &= \pi\gamma^{-1}(1-x^2-y^2)^{1-\gamma}, \qquad (x,y) \in B_1(0), \\ w_3^\gamma(x,y) &= \pi^{-1}(\gamma+1)(1-x^2-y^2)^{\gamma}, \qquad (x,y) \in B_1(0). \end{aligned}$$

Assume that $\lambda = 1$, in which case the singular functions in detector space, for arbitrary γ, are given by

$$\Phi^\gamma_{m,l}(\varphi,s) = \frac{C_m^\gamma(s)e^{-i(m-2l)\varphi}}{\sqrt{\pi d_m^\gamma w_1^\gamma(s)}},$$

and the singular values by $\sigma_{m,l} = \sqrt{\pi c^\gamma_{m,l}}$, where

$$d_m^\gamma = \frac{\sqrt{\pi}\gamma 2^{1-2\gamma}\Gamma(m+2\gamma)}{m!\Gamma(\gamma+1/2)(m+\gamma)\Gamma(\gamma)}, \qquad c^\gamma_{m,l} = \binom{m}{l}\frac{\Gamma(2\gamma)\Gamma(\gamma+m-l)\Gamma(\gamma+l)}{\Gamma(2\gamma+m)\Gamma(\gamma)^2}.$$

Hence, in view of Lemma 4.8, the eigenfunctions in brain space can be written as

$$\Psi^{\gamma}_{m,l}(x,y) = \frac{1}{\pi\sqrt{d^{\gamma}_m c^{\gamma}_{m,l}}\, w^{\gamma}_2(x,y)} \int_{-\pi/2}^{\pi/2} C^{\gamma}_m(x\cos\varphi + y\sin\varphi) e^{-i(m-2l)\varphi}\, d\varphi$$
$$= \frac{h^{\gamma}_{m,l}(x,y)}{\pi\sqrt{d^{\gamma}_m c^{\gamma}_{m,l}}\, w^{\gamma}_2(x,y)}$$

with $h^{\gamma}_{m,l}$ defined in (4.49), and in particular, regarding (4.47) and minding that $d^1_m = 1$, $c^1_{m,l} = (m+1)^{-1}$, and $w^1_2(x,y) = \pi$, the Zernike functions are given by

$$z_{m,l}(x,y) = \frac{\pi}{\sqrt{m+1}} \Psi^1_{m,l}(x,y) = \pi^{-1} h^1_{m,l}(x,y). \tag{4.51}$$

We now come back to the differential operators $D = (\partial/\partial x - i\partial/\partial y)/2$ and $\bar{D} = (\partial/\partial x + i\partial/\partial y)/2$. From the Gegenbauer identity $d/ds C^{\gamma}_m(s) = 2\gamma C^{\gamma+1}_{m-1}(s)$, see e. g. (4.7.14) in Szegö (1967), it readily follows that

$$Dh^{\gamma}_{m,l} = \gamma h^{\gamma+1}_{m-1,l-1}, \qquad \bar{D}h^{\gamma}_{m,l} = \gamma h^{\gamma+1}_{m-1,l},$$

where in particular $Dh^{\gamma}_{m,0} = \bar{D}h^{\gamma}_{m,m} = 0$. For $p,q \in \mathbb{N}$ such that $p+q = s$ this provides (4.48).

The norm of these derivatives can now be evaluated with respect to μ^s_3. For this, note that $w^{\gamma}_3 = (w^{\gamma+1}_2)^{-1}$ and that the $\Psi^{\gamma}_{m,l}$ are normalized with respect to μ^{γ}_2. Therefore,

$$\|h^{\gamma+1}_{m,l}\|_{\mu^{\gamma}_3} = \pi\sqrt{d^{\gamma+1}_m c^{\gamma+1}_{m,l}}\, \|w^{\gamma+1}_2 \Psi^{\gamma+1}_{m,l}\|_{\mu^{\gamma}_3} = \pi\sqrt{d^{\gamma+1}_m c^{\gamma+1}_{m,l}}\, \|\Psi^{\gamma+1}_{m,l}\|_{\mu^{\gamma+1}_2}$$
$$= \pi\sqrt{d^{\gamma+1}_m c^{\gamma+1}_{m,l}},$$

for $p,q \in \mathbb{N}$ such that $p+q = s$ yielding

$$\|D^p \bar{D}^q z_{m,l}\|_{\mu^s_3} = \frac{s!}{\pi} \|h^{s+1}_{m-s,l-p}\|_{\mu^s_3} = s!\sqrt{d^{\gamma+1}_{m-s} c^{\gamma+1}_{m-s,l-p}}.$$

Plugging in the formulas for $c^{\gamma}_{m,l}$ and d^{γ}_m given above provides (4.50). □

4.6.4. Proofs of some technical lemmas

Lemma 4.4. For all $\gamma > 1$ and $\delta, c_1, c_2 \in \mathbb{R}$,

$$\sum_{j=0}^{n} \gamma^{-j}(n-j)^{c_1}(j+\delta)^{c_2} \sim n^{c_1} \sum_{j=0}^{\infty} \gamma^{-j}(j+\delta)^{c_2} \qquad \text{as } n \to \infty.$$

Proof. Assume that $c_1 \geq 0$, the case $c_1 < 0$ is analogous. Then, for all n,

$$\sum_{j=0}^{n} \gamma^{-j}(1-j/n)^{c_1}(j+\delta)^{c_2} \leq \sum_{j=0}^{\infty} \gamma^{-j}(j+\delta)^{c_2},$$

providing the upper bound. To establish the lower bound, let $0<\varepsilon<1$ and set $n_\varepsilon=\lfloor \varepsilon n\rfloor$. Then,

$$\sum_{j=n_\varepsilon+1}^{n}\gamma^{-j}(1-j/n)^{c_1}(j+\delta)^{c_2}\le(1-\varepsilon)^{c_1}\sum_{j=n_\varepsilon+1}^{n}\gamma^{-j}(j+\delta)^{c_2}\longrightarrow 0,$$

so that

$$\begin{aligned}\lim_{n\to\infty}\sum_{j=0}^{n}\gamma^{-j}(1-j/n)^{c_1}(j+\delta)^{c_2}&\ge(1-\varepsilon)^{c_1}\lim_{n\to\infty}\sum_{j=0}^{n_\varepsilon}\gamma^{-j}(j+\delta)^{c_2}\\&=(1-\varepsilon)^{c_1}\sum_{j=0}^{\infty}\gamma^{-j}(j+\delta)^{c_2}.\end{aligned}$$

Now letting $\varepsilon\to 0$ provides the lower bound and concludes the proof. □

Lemma 4.6. If there exist $\beta\ge 1$, $\delta>0$, and a positive, bounded sequence $c=(c_0,c_1,\dots)$ such that

$$\alpha_{j+k,k}^{-1}=c_jk^\beta+O\big(((j+1)(k+1))^{\beta-\delta}\big),\qquad j,k\ge 0,\tag{4.52}$$

then, for all $\alpha\ge 0$,

$$\sum_{(j,k)\in(n)}(j+k+1)(j+1)^\alpha(k+1)^\alpha\alpha_{j+k,k}^{-1}\sim\frac{K(\beta+1,c)}{\alpha+\beta+2}n^{\alpha+\beta+2}$$

as $n\to\infty$, where $K(\beta,c)=\sum_{j=0}^{\infty}c_j(j+1)^{-(\beta+1)}$.

Proof. Conveniently, assume that $\delta\le 1$, and set $[n]=\{(j,k):j,k\ge 1,jk\le n\}$ as well as $\bar\alpha_{j+k,k}=\alpha_{j+k-2,k-1}$ and $\bar c_j=c_{j-1}$, so that the sum above reads

$$\begin{aligned}&\sum_{(j,k)\in[n]}(j+k-1)j^\alpha k^\alpha\bar\alpha_{j+k,k}^{-1}\\&=\sum_{(j,k)\in[n]}j^\alpha k^{\alpha+1}\bar\alpha_{j+k,k}^{-1}+\sum_{(j,k)\in[n]}j^{\alpha+1}k^\alpha\bar\alpha_{j+k,k}^{-1}-\sum_{(j,k)\in[n]}j^\alpha k^\alpha\bar\alpha_{j+k,k}^{-1}.\end{aligned}$$

Denote these latter three sums by $S_{1,n}$, $S_{2,n}$, and $S_{3,n}$, respectively. We will see that the first sum $S_{1,n}$ is the driving part. In fact, $S_{3,n}$ is bounded by $S_{2,n}$ which itself will be shown to be negligible at rate $n^{\alpha+\beta+2}$.

Remember the approximation

$$\sum_{j=1}^{\lfloor x\rfloor}j^\gamma=(\gamma+1)^{-1}x^{\gamma+1}+O\big(x^\gamma\big)=O\big(x^{\gamma+1}\big),\quad x\ge 1,\gamma\ge 0,$$

where the constants hidden in the O-terms only depend on γ, no longer on x. Further, using $|k^\beta-(k-1)^\beta|=O(k^{\beta-1})$ and the boundedness of the c_j, (4.52) gives $\bar\alpha_{j+k,k}=\bar c_jk^\beta+O\big(((j+

$1)(k+1))^{\beta-\delta}\big)$, so for any $x \geq 1, \gamma \geq 0$, and $j \in \mathbb{N}$,

$$\begin{aligned}\sum_{k=1}^{\lfloor x\rfloor} k^\gamma \bar{\alpha}_{j+k,k}^{-1} &= \bar{c}_j \sum_{k=1}^{\lfloor x\rfloor} k^{\gamma+\beta} + \sum_{k=1}^{\lfloor x\rfloor} O\big(j^{\beta-\delta}k^{\gamma+\beta-\delta}\big)\\ &= \frac{\bar{c}_j}{\gamma+\beta+1}x^{\gamma+\beta+1} + O\big(j^{\beta-\delta}x^{\gamma+\beta+1-\delta}\big).\end{aligned}$$

The sum $S_{2,n}$ therefore satisfies

$$\begin{aligned}S_{2,n} &= \sum_{j=1}^{n} j^{\alpha+1} \sum_{k=1}^{\lfloor n/j\rfloor} k^\alpha \bar{\alpha}_{j+k,k}^{-1}\\ &= \sum_{j=1}^{n} j^{\alpha+1}\big(O\big((n/j)^{\alpha+\beta+1}\big) + O\big(j^{\beta-\delta}(n/j)^{\alpha+\beta+1-\delta}\big)\big)\\ &= n^{\alpha+\beta+1}\sum_{j=1}^{n} O\big(j^{-\beta}\big) + n^{\alpha+\beta+1-\delta}\sum_{j=1}^{n} O(1)\\ &= O\big(n^{\alpha+\beta+1}\log n\big) + O\big(n^{\alpha+\beta+2-\delta}\big),\end{aligned}$$

providing the negligibility of $S_{2,n}$ and $S_{3,n}$. Finally, the first sum $S_{1,n}$ gives

$$\begin{aligned}S_{1,n} &= \sum_{j=1}^{n} j^{\alpha} \sum_{k=1}^{\lfloor n/j\rfloor} k^{\alpha+1} \bar{\alpha}_{j+k,k}^{-1}\\ &= \sum_{j=1}^{n} \frac{\bar{c}_j j^\alpha}{\alpha+\beta+2}(n/j)^{\alpha+\beta+2} + \sum_{j=1}^{n} j^\alpha O\big(j^{\beta-\delta}(n/j)^{\alpha+\beta+2-\delta}\big)\\ &= \frac{n^{\alpha+\beta+2}}{\alpha+\beta+2}\sum_{j=1}^{n} \bar{c}_j j^{-(\beta+2)} + n^{\alpha+\beta+2-\delta}\sum_{j=1}^{n} O\big(j^{-2}\big)\\ &= \frac{K(\beta+1,c)n^{\alpha+\beta+2}}{\alpha+\beta+2}\big(1+o(1)\big) + O\big(n^{\alpha+\beta+2-\delta}\big),\end{aligned}$$

which concludes the proof. □

4.6.5. Exact rates for the ordinary Radon transform

To complement the above analysis, we finally show that in contrast to the weight function λ on the angle, which strongly effects the rate of convergence, the parameter γ in the weight functions w_1 and w_2 alone does not influence the rate of convergence.

In case that $\lambda = 1$, i. e., the Radon transform inverse problem as studied in the past, exact minimax rates can be given not only for $\gamma = 1$, the situation for which the rates are well known, but for arbitrary γ. We here concentrate on the case $\gamma \in (0,1]$, including parallel beam design, for instance.

Recall that for $\lambda = 1$ the singular values $\sigma_{m,l}$ are given by

$$\sigma_{m,l} = \sqrt{\pi c_{m,l}}$$

with

$$c_{m,l} = \binom{m}{l} \frac{\Gamma(2\gamma)\Gamma(l+\gamma)\Gamma(m-l+\gamma)}{\Gamma(m+2\gamma)\Gamma(\gamma)^2}.$$

In view of Lemma 4.12 and using $\Gamma(m+2\gamma)/\Gamma(m+1) \sim m^{2\gamma-1}$,

$$\sum_{l=0}^{m} c_{m,l}^{-1} = \frac{\Gamma(\gamma)^2}{\Gamma(2\gamma)} \frac{\Gamma(m+2\gamma)}{\Gamma(m+1)} \sum_{l=0}^{m} \frac{\Gamma(l+1)}{\Gamma(l+\gamma)} \frac{\Gamma(m-l+1)}{\Gamma(m-l+\gamma)}$$
$$\sim C_\gamma m^2,$$

as $m \to \infty$, where $C_\gamma = \frac{\sqrt{\pi}\Gamma(\gamma)^2\Gamma(2-\gamma)}{\Gamma(2\gamma)\Gamma(5/2-\gamma)2^{3-2\gamma}}$. Since this can be treated as imposing A 4.2 for $\rho = 2$ and $C = C_\gamma$, Theorem 4.4 provides the minimax risk

$$r_\varepsilon(\Theta_1(\kappa, L)) \sim C_1^* \varepsilon^{\frac{4\kappa}{2\kappa+3}} \qquad \text{as } \varepsilon \to 0$$

with

$$C_1^* = \Big(\frac{C_\gamma \kappa}{\pi(\kappa+3)}\Big)^{\frac{2\kappa}{2\kappa+3}} \frac{\big(L(2\kappa+3)\big)^{\frac{3}{2\kappa+3}}}{3}.$$

For example, using the duplication formula $\Gamma(z)\Gamma(z+0.5) = 2^{1-2z}\sqrt{\pi}\Gamma(2z)$, $z \in \mathbb{Z}$, in parallel beam design we particularly have

$$C_{0.5} = \pi^2/8.$$

Lemma 4.12. *Denoting by Γ the Gamma function, for any $\gamma \in (0,1]$,*

$$\sum_{l=0}^{m} \frac{\Gamma(l+1)}{\Gamma(l+\gamma)} \frac{\Gamma(m-l+1)}{\Gamma(m-l+\gamma)} \sim \frac{\sqrt{\pi}\Gamma(2-\gamma)}{\Gamma(5/2-\gamma)} 2^{2\gamma-3} m^{3-2\gamma} \qquad \textit{as } m \to \infty.$$

Proof. Set $f(x) = \Gamma(x)/\Gamma(x+\gamma-1)$, and without loss of generality always assume that m is even. Then, by symmetrie in l and $m-l$,

$$\sum_{l=0}^{m} \frac{\Gamma(l+1)}{\Gamma(l+\gamma)} \frac{\Gamma(m-l+1)}{\Gamma(m-l+\gamma)} = 2\sum_{l=0}^{m/2} f(l+1)f(m-l+1).$$

Let $\varepsilon > 0$. As $x \to \infty$, the function f satisfies $f(x) \sim x^{1-\gamma}$, whence there exists $x_\varepsilon > 0$ such that

$$1-\varepsilon \leq f(x+1)/x^{1-\gamma} \leq 1+\varepsilon, \qquad x \geq x_\varepsilon. \tag{4.53}$$

Setting $m_\varepsilon = \lceil x_\varepsilon \rceil$, it is evident that

$$\sum_{l=0}^{m_\varepsilon - 1} f(l+1)f(m-l+1) = O(m^{1-\gamma}).$$

Further,

$$\sum_{l=m_0}^{m/2} f(l+1)f(m-l+1) \gtrsim \sum_{l=m_0}^{m/2} l^{1-\gamma}(m-l)^{1-\gamma} \geq (m/2)^{1-\gamma} \sum_{l=m_0}^{m/2} l^{1-\gamma}$$
$$\gtrsim m^{3-2\gamma}.$$

For each ε fixed, we therefore obtain the upper bound

$$\limsup_{m\to\infty}\sum_{l=0}^{m/2} f(l+1)f(m-l+1) \le ((1+\varepsilon)^2+o(1))\limsup_{m\to\infty}\sum_{l=0}^{m/2} l^{1-\gamma}(m-l)^{1-\gamma},$$

and likewise the lower bound

$$\liminf_{m\to\infty}\sum_{l=0}^{m/2} f(l+1)f(m-l+1) \ge ((1-\varepsilon)^2+o(1))\liminf_{m\to\infty}\sum_{l=0}^{m/2} l^{1-\gamma}(m-l)^{1-\gamma},$$

so letting $\varepsilon \to 0$ gives

$$\sum_{l=0}^{m/2} f(l+1)f(m-l+1) \sim \sum_{l=0}^{m/2} l^{1-\gamma}(m-l)^{1-\gamma} = m^{3-2\gamma}\frac{1}{m}\sum_{l=0}^{m/2}(l/m)^{1-\gamma}(1-l/m)^{1-\gamma}$$
$$\sim m^{3-2\gamma}\int_0^{1/2} x^{1-\gamma}(1-x)^{1-\gamma}dx.$$

With this, and minding that

$$\int_0^1 x^{1-\gamma}(1-x)^{1-\gamma}dx = \frac{\sqrt{\pi}\Gamma(2-\gamma)2^{2\gamma-3}}{\Gamma(5/2-\gamma)},$$

we conclude the proof. □

Part III

Supplement

5

Two-component mixtures with independent coordinates as conditional mixtures: Technical report

5.1. Identification

Consider the conditional mixture

$$F(y|B) = (1-\pi(B))F_0(y) + \pi(B)F_1(y), \qquad y \in \mathbb{R}, B \in \mathscr{B}^p, \tag{5.1}$$

with mixture weight function π and component distribution functions F_0 and F_1. To identify the components in (5.1), assume that

Assumption A 5.1. 1. there exist $B_0, B_1 \in \mathscr{B}^p$ such that $0 < \pi(B_0), \pi(B_1) < 1$, $\pi(B_0) \neq \pi(B_1)$. 2. there exists a $y_0 \in \mathbb{R}$ such that $F_1(y_0) \neq F_0(y_0)$.

Further, consider the following tail conditions.

C1. $\lim_{y\to-\infty} F_1(y)/F_0(y) = 0$

C2. $\lim_{y\to+\infty}(1-F_0(y))/(1-F_1(y)) = 0$

C3. $\lim_{y\to+\infty} \tilde{F}_0(y)/\tilde{F}_1(y) = 0$

Hohmann and Holzmann (2013b, Theorem 3.1) show that mixture (5.1) is identifiable under A 5.1 and assuming C1 and C2/C3. The following example shows that is does not suffice to impose only one of the tail conditions above.

Example 5.1. Assume that mixture (5.1) is identifiable in the sense of Theorem 3.1 in Hohmann and Holzmann (2013b). Let $\pi_2 : \mathscr{B}^p \to [0,1]$ be a different weight function such that $\pi_2(B_0) < \pi_1(B_0)$, and set

$$\pi_2(B) = 1 - \frac{(1-\pi_1(B))(1-\pi_2(B_0))}{1-\pi_1(B_0)}, \qquad B \in \mathscr{B}\backslash\{B_0\}. \tag{5.2}$$

Further, set $G_1 = F_1$, and define G_0 according to

$$G_0(y) = \frac{1-\pi_1(B_0)}{1-\pi_2(B_0)} F_0(y) + \Big(1 - \frac{1-\pi_1(B_0)}{1-\pi_2(B_0)}\Big) F_1(y).$$

Then, G_0 is indeed a distribution function due to $\pi_1(B_0) > \pi_2(B_0)$, and by construction the ratio $\rho = (1-\pi_1(B))/(1-\pi_2(B))$ does not depend on B, so for all $y \in \mathbb{R}$ and $B \in \mathscr{B}^p$ we obtain that

$$G(y|B) = (1-\pi_2(B))G_0(y) + \pi_2(B)G_1(y) = F(y|B).$$

Also, G_0 and G_1 meet C1 since

$$\frac{G_1(y)}{G_0(y)} = \frac{F_1(y)/F_0(y)}{\rho + (1-\rho)F_1(y)/F_0(y)} \longrightarrow 0 \qquad \text{as } y \to -\infty,$$

while, in general, they satisfy neither of the conditions C2 and C3. □

The next example shows that also the role of the conditioning events B_0 and B_1 is important for the nonparametric identification of a two-component mixture. In fact, even with known mixture proportion π, regularity conditions such as C1 and C2 do not provide the identification of an ordinary mixture

$$F(y) = (1-\pi)F_0(y) + \pi F_1(y). \tag{5.3}$$

Example 5.2. Assume that F_0 and F_1 in (5.3) are absolutely continuous with densities f_0 and f_1, respectively, and assume that there exist $a, b \in \mathbb{R}$, $a < b$, such that F_0 is strictly concave on the interval (a,b) and

$$f_0(y) + \frac{\pi}{1-\pi} f_1(y) \geq \frac{F_0(b) - F_0(a)}{b-a}, \qquad a \leq y \leq b. \tag{5.4}$$

Set $G_0 = F_0 \mathbf{1}_{[a,b)^\complement} + F_a^b$ and $G_1 = F_1 + \frac{1-\pi}{\pi}(F_0 \mathbf{1}_{[a,b)} - F_a^b)$, where

$$F_a^b(y) = \Big(\frac{(y-a)F_0(b) - (y-b)F_0(a)}{b-a}\Big)\mathbf{1}_{[a,b)}(y).$$

(5.4) guarantees that G_1 is non-decreasing and thus a distribution function. Now G_0 and G_1 adopt C1 and C2 from F_0 and F_1, and the mixture $G(y) = (1-\pi)G_0 + \pi G_1(y)$ satisfies $G = F$. □

5.2. Estimating quotients in the tails

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be mutually independent sequences of i.i.d. observations with distribution functions F and G, respectively, and assume that

$$F(y)/G(y) \longrightarrow \theta \qquad \text{as } y \to -\infty \tag{5.5}$$

and

$$\widetilde{F}(y), \widetilde{G}(y) \longrightarrow 0, \quad \widetilde{F}(y)/\widetilde{G}(y) \longrightarrow \eta \qquad \text{as } y \to \infty \tag{5.6}$$

hold for some $\theta > 0$ and $\eta \in \mathbb{C}\backslash\{0\}$, where $\widetilde{F}$ and $\widetilde{G}$ denote the characteristic functions of F and G. We shall construct asymptotically normal estimators of θ and η. In the following, suppose that l_n and m_n are sequences in $\mathbb{N}$ such that $l_n, m_n \asymp n$ as $n \to \infty$.

To estimate η in (5.6), let

$$\eta_n = \widetilde{F}_n(h_n)/\widetilde{G}_n(h_n), \qquad \widetilde{F}_n(y) = \frac{1}{l_n}\sum_{k=1}^{l_n} \exp(iyX_k), \qquad \widetilde{G}_n(y) = \frac{1}{m_n}\sum_{k=1}^{m_n} \exp(iyY_k),$$

with h_n a sequence tending to infinity. Decompose

$$\eta_n - \eta = (\eta_n - \bar{\eta}_n) + (\bar{\eta}_n - \eta), \qquad \bar{\eta}_n = \widetilde{F}(h_n)/\widetilde{G}(h_n).$$

In order to handle the "variance term", write

$$\sqrt{r_n}(\eta_n - \bar{\eta}_n) = \frac{\sqrt{r_n/m_n}}{\widetilde{G}_n(h_n)}\Big(\sqrt{m_n/l_n}\,\widetilde{\mathbf{F}}_n(h_n) - \bar{\eta}_n\widetilde{\mathbf{G}}_n(h_n)\Big), \tag{5.7}$$

where $\widetilde{\mathbf{F}}_n = \sqrt{l_n}(\widetilde{F}_n - \widetilde{F})$ and $\widetilde{\mathbf{G}}_n = \sqrt{m_n}(\widetilde{G}_n - \widetilde{G})$ are the characteristic processes and $r_n \to \infty$. Assume that r_n satisfies

$$r_n/n \to 0, \quad r_n/\sqrt{n} \to \infty \qquad \text{as } n \to \infty, \tag{5.8}$$

and that $h_n \to_p \infty$ is chosen such that

$$|\widetilde{G}_n(h_n)| = \sqrt{r_n/m_n}(1 + o_P(1)). \tag{5.9}$$

We shall use strong approximations of the characteristic processes by

$$\mathbf{C}(y) = \int \exp(iyx)\,\mathbf{B}(F(dx)) \tag{5.10}$$

for $\widetilde{\mathbf{F}}_n$, and similarly for $\widetilde{\mathbf{G}}_n$. In order that these processes are sample-continuous and that strong approximations work, some conditions on F and G are required, see Csörgő (1981). We shall adopt the following sufficient condition: Assume that there exists $\gamma > 0$ such that

$$y^{\gamma}H(-y) + y^{\gamma}(1 - H(y)) = O(1) \quad \text{as } y \to \infty, \quad H = F, G. \tag{5.11}$$

Finally, we assume that there also exists a non-random sequences $t_n \to \infty$ such that

$$t_n = o\Big(n^{\gamma/(2\gamma+4)}(\log n)^{-(\gamma+1)/(\gamma+2)}\Big), \tag{5.12}$$
$$|h_n - t_n| = o_P(1), \tag{5.13}$$
$$|\widetilde{G}(h_n) - \widetilde{G}(t_n)| = o_P\Big(\sqrt{r_n/m_n}\Big), \tag{5.14}$$

with γ determined by (5.11).

To estimate θ in (5.5), let

$$\theta_n = F_n(h_n)/G_n(h_n), \qquad F_n(y) = \frac{1}{l_n}\sum_{k=1}^{l_n}\mathbf{1}_{\{X_k \le y\}}, \quad G_n(y) = \frac{1}{m_n}\sum_{k=1}^{m_n}\mathbf{1}_{\{Y_k \le y\}}, \tag{5.15}$$

where the level h_n is specified below. Write

$$\theta_n - \theta = (\theta_n - \bar{\theta}_n) + (\bar{\theta}_n - \theta), \qquad \bar{\theta}_n = F(h_n)/G(h_n).$$

Again assume that $r_n \to \infty$ satisfies (5.8), and that $h_n \to_p -\infty$ is chosen such that

$$G_n(h_n) = r_n/m_n + o_P(r_n/n) = r_n/m_n\,(1 + o_P(1)). \tag{5.16}$$

(5.16) is satisfied if we choose in particular $h_n = Y_{m_n(\lfloor r_n \rfloor)}$, where $\lfloor r_n \rfloor$ is the largest integer smaller than r_n, and where $Y_{m_n(\lfloor r_n \rfloor)}$ denotes the $\lfloor r_n \rfloor$-th largest order statistic of the sample $Y_1, \ldots, Y_{m_n}$, since $G_n(h_n) = \lfloor r_n \rfloor / m_n = r_n/m_n(1+o(1))$.

Theorem 5.1. *Suppose that (5.8), (5.9) and (5.11)-(5.16) hold. If there exists $\tau > 0$ such that $m_n/l_n \to \tau$, then*

$$\sqrt{r_n}\begin{pmatrix}\theta_n - \bar{\theta}_n \\ \mathrm{Re}\left(\eta_n - \bar{\eta}_n\right) \\ \mathrm{Im}\left(\eta_n - \bar{\eta}_n\right)\end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \frac{1}{2}\begin{pmatrix}2(\tau\theta+\theta^2) & 0 & 0 \\ 0 & \tau+|\eta|^2 & 0 \\ 0 & 0 & \tau+|\eta|^2\end{pmatrix}\right).$$

The proof of Theorem 5.1 proceeds in several steps. The asymptotic normality of $\sqrt{r_n}(\eta_n - \bar{\eta}_n)$ was shown in Hohmann and Holzmann (2013b). We continue by showing that

$$\sqrt{r_n}\left(\theta_n - \bar{\theta}_n\right) \rightsquigarrow \mathcal{N}\left(0, \tau\theta + \theta^2\right), \tag{5.17}$$

for which we need the following additional results.

Lemma 5.1. *Let $r_n \to \infty$, $r_n/n \to 0$ as $n \to \infty$, and $\log n/\sqrt{r_n} \to 0$. Then on a sufficiently rich probability space there exist versions of the X_k and Y_k, and independent sequences $\mathbf{B}_{1,n}$ and $\mathbf{B}_{2,n}$ of standard Brownian bridges on $[0,1]$ such that*

$$\left\|\mathbf{F}_n - \mathbf{B}_{1,l_n} \circ F\right\|_\infty = o_P\left(\sqrt{r_n/n}\right), \quad \left\|\mathbf{G}_n - \mathbf{B}_{2,m_n} \circ G\right\|_\infty = o_P\left(\sqrt{r_n/n}\right).$$

See del Barrio, Deheuvels and van de Geer (2007).

Lemma 5.2. *Let l_n be a sequence in $\mathbb{N}$, $\mathbf{B}_n$ be a sequence of standard Brownian bridges on $[0,1]$, and X_n be a random sequence (not necessarily independent of $\mathbf{B}_n$) such that $X_n \to_P \gamma$ for some $\gamma \geq 0$. For all real $c_n \downarrow 0$ it holds that*

$$\left|\mathbf{B}_{l_n}(c_n X_n) - \mathbf{B}_{l_n}(c_n \gamma)\right| = o_P(\sqrt{c_n}).$$

Proof. Let $Z_1, Z_2, \ldots$ be a sequence of standard normal variables, and for $n \in \mathbb{N}$ and $t \in [0,1]$ define $\mathbf{W}_n(t) = \mathbf{B}_n(t) + tZ_n$. Then $\mathbf{W}_n$ is a sequence of standard Wiener processes, and

$$\begin{aligned}\mathbf{B}_{l_n}(c_n X_n) - \mathbf{B}_{l_n}(c_n\gamma) &= \mathbf{W}_{l_n}(c_n X_n) - \mathbf{W}_{l_n}(c_n\gamma) + c_n(\gamma - X_n)Z_{l_n} \\ &= \mathbf{W}_{l_n}(c_n X_n) - \mathbf{W}_{l_n}(c_n\gamma) + o_P(c_n),\end{aligned}$$

so that the limit behavior under consideration de facto only depends on the properties of Brownian motion. For all $\varepsilon, \delta > 0$,

$$\begin{aligned}&\mathrm{P}\left(c_n^{-1/2}|\mathbf{W}_{l_n}(c_n X_n) - \mathbf{W}_{l_n}(c_n\gamma)| > \varepsilon\right) \\ &\quad \leq \mathrm{P}(|X_n - \gamma| > \delta) + \mathrm{P}\left(\sup_{|t-\gamma|\leq\delta} c_n^{-1/2}|\mathbf{W}_{l_n}(c_n t) - \mathbf{W}_{l_n}(c_n\gamma)| > \varepsilon\right) \\ &\quad = \mathrm{P}(|X_n - \gamma| > \delta) + \mathrm{P}\left(\sup_{|t-\gamma|\leq\delta} |\mathbf{W}_1(t) - \mathbf{W}_1(\gamma)| > \varepsilon\right)\end{aligned}$$

since, by Brownian scaling, each process $y \mapsto c_n^{-1/2}\mathbf{W}_{l_n}(c_n y)$ is itself a standard Brownian motion. Note also that, by the continuity of Brownian motion sample paths, the supremum has to be taken over $t \in \mathbb{Q}$ only, what makes it a measurable function. The left probability tends to zero for all $\delta > 0$ because $X_n \to_P \gamma$. The right probability can be made arbitrarily small by the choice of δ since, again by the almost sure continuity of $\mathbf{W}_1$,

$$\lim_{m\to\infty} \mathrm{P}\Big(\sup_{|t-\gamma|\leq 1/m} |\mathbf{W}_1(t) - \mathbf{W}_1(\gamma)| > \varepsilon \Big) = \mathrm{P}\Big(\bigcap_{m\in\mathbb{N}} \Big\{ \sup_{|t-\gamma|\leq 1/m} |\mathbf{W}_1(t) - \mathbf{W}_1(\gamma)| > \varepsilon \Big\} \Big) = 0\,.$$

Conclude that $c_n^{-1/2}|\mathbf{B}_{l_n}(c_n X_n) - \mathbf{B}_{l_n}(c_n\gamma)| \to_P 0$. □

Since $\mathbf{B}$ is zero mean Gaussian with covariance $\mathrm{E}\big(\mathbf{B}(s)\mathbf{B}(t)\big) = (s\wedge t) - st$, $s,t \in [0,1]$, it readily follows that

$$c_n^{-1/2}\mathbf{B}(c_n\gamma) \sim_d \mathcal{N}\big(0, \gamma(1-c_n\gamma)\big) \longrightarrow \mathcal{N}(0,\gamma) \quad \text{as } n\to\infty\,. \tag{5.18}$$

Proof of (5.17). Write

$$\sqrt{r_n}\big(\theta_n - \bar\theta_n\big) = \frac{\sqrt{r_n/m_n}}{G_n(h_n)}\Big(\sqrt{m_n/l_n}\,\mathbf{F}_n(h_n) - \bar\theta_n\,\mathbf{G}_n(h_n)\Big), \tag{5.19}$$

where $\mathbf{F}_n = \sqrt{l_n}(F_n - F)$ and $\mathbf{G}_n = \sqrt{m_n}(G_n - G)$ denote the empirical processes. By Lemma 5.1 there exists a sequence $\mathbf{B}_{2,n}$ of standard Brownian bridges such that

$$\mathbf{G}_n(h_n) = \mathbf{B}_{2,m_n}(G(h_n)) + o_P\big(\sqrt{r_n/n}\big). \tag{5.20}$$

Now, (5.8) and (5.16) imply that

$$\frac{n}{r_n}\big|G(h_n) - r_n/m_n\big| \leq \frac{n}{r_n}\|G - G_n\|_\infty + o_P(1) = \frac{\sqrt{n}}{r_n}O_P(1) + o_P(1) = o_P(1),$$

yielding $G(h_n) = r_n/m_n\,\big(1 + o_P(1)\big)$. Inserting this in (5.20) and using Lemma 5.2 (with $\gamma = 1$) we find that

$$\mathbf{G}_n(h_n) = \mathbf{B}_{2,m_n}(r_n/m_n) + o_P\big(\sqrt{r_n/n}\big).$$

Similarly for $\mathbf{F}_n(h_n)$, there is an independent sequence $\mathbf{B}_{1,n}$ of standard Brownian bridges such that, using $\bar\theta_n \to_P \theta$ and $F(h_n) = \bar\theta_n G(h_n) = r_n/m_n\big(\bar\theta_n + o_P(1)\big)$,

$$\mathbf{F}_n(h_n) = \mathbf{B}_{1,l_n}(\theta\, r_n/m_n) + o_P\big(\sqrt{r_n/n}\big).$$

Therefore, using (5.19) and (5.16),

$$\begin{aligned}\sqrt{r_n}\big(\theta_n - \bar\theta_n\big) &= \frac{\sqrt{r_n/m_n}}{G_n(h_n)}\Big(\sqrt{\frac{m_n}{l_n}}\mathbf{F}_n(h_n) - \bar\theta_n\mathbf{G}_n(h_n)\Big)\\ &= \frac{\sqrt{m_n/r_n}}{1+o_P(1)}\Big(\sqrt{\tau}\,\mathbf{B}_{1,l_n}(\theta\, r_n/m_n) - \theta\,\mathbf{B}_{2,m_n}(r_n/m_n)\Big) + o_P(1)\,,\end{aligned} \tag{5.21}$$

so that the result follows from (5.18) and the independence of $\mathbf{B}_{1,n}$ and $\mathbf{B}_{2,n}$. □

Proof of asymptotic independence in Theorem 5.1. We say that sequences of random vectors X_n in $\mathbb{R}^p$ and Y_n in $\mathbb{R}^q$ are asymptotically independent if

$$\mathrm{E}\big(f(X_n)g(Y_n)\big) - \mathrm{E}\big(f(X_n)\big)\mathrm{E}\big(g(Y_n)\big) \longrightarrow 0 \quad \text{as } n\to\infty$$

for all bounded, non-negative, Lipschitz functions f and g on $\mathbb{R}^p$ and $\mathbb{R}^q$, resp. For the next lemma see Example 1.4.6 in van der Vaart and Wellner (2000).

Lemma 5.3. *If there exist independent random vectors X and Y such that $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$, and if further X_n and Y_n are asymptotically independent, then $(X_n, Y_n)' \rightsquigarrow (X, Y)'$.*

The following lemma gives a criterion for asymptotic independence in case of Gaussian sequences, where boundedness and convergence of matrices is understood with respect to the Frobenius norm given by $\|M\|_F = \big(\sum_{i,j} |M_{i,j}|^2\big)^{1/2}$.

Lemma 5.4. *If X_n and Y_n are zero mean jointly Gaussian, the covariance matrices $\operatorname{Cov} X_n$ and $\operatorname{Cov} Y_n$ are uniformly bounded above and uniformly bounded away from zero, eventually, and if $\operatorname{Cov}(X_n, Y_n) \to 0$, then X_n and Y_n are asymptotically independent.*

Proof. Let f and g be positive, bounded, and Lipschitz. Denoting by $\phi_{(X_n,Y_n)}$ the joint density and by ϕ_{X_n} and ϕ_{Y_n} the marginal densities of X_n and Y_n,

$$\big|\mathrm{E}\big(f(X_n)g(Y_n)\big) - \mathrm{E}\big(f(X_n)\big)\mathrm{E}\big(g(Y_n)\big)\big| \le \iint f(x)g(y)\big|\phi_{(X_n,Y_n)}(x,y) - \phi_{X_n}(x)\phi_{Y_n}(y)\big|\,dx\,dy\,.$$

By the boundedness of $\operatorname{Cov} X_n$ and $\operatorname{Cov} Y_n$, and by the convergence $\operatorname{Cov}(X_n, Y_n) \to 0$,

$$|\phi_{(X_n,Y_n)}(x,y) - \phi_{X_n}(x)\phi_{Y_n}(y)| \longrightarrow 0\,, \quad x, y \in \mathbb{R}\,.$$

Hence, minding that the densities are uniformly bounded above by an integrable function due to the boundedness of $\operatorname{Cov} X_n$ and $\operatorname{Cov} Y_n$, the result follows in view of Lebesgue's dominated convergence. □

We are now ready to come back to the estimators θ_n and η_n. Regarding (5.21), (5.7) and Lemma 3.3 in Hohmann and Holzmann (2013b), under certain assumptions there exist independent sequences $\mathbf{B}_{1,n}$ and $\mathbf{B}_{2,n}$ of standard Brownian bridges such that

$$\begin{aligned}
\sqrt{r_n}\big(\theta_n - \bar\theta_n\big) &= \frac{\sqrt{m_n/r_n}}{1+o_P(1)}\Big(\sqrt{\tau}\,\mathbf{B}_{1,l_n}(\theta\, r_n/m_n) - \theta\,\mathbf{B}_{2,m_n}(r_n/m_n)\Big) + o_P(1),\\
\sqrt{r_n}\big(\eta_n - \bar\eta_n\big) &= \frac{z_n}{(1+o_P(1))}\Big(\sqrt{\tau}\,\mathbf{C}_{1,l_n}(t_n) - \eta\,\mathbf{C}_{2,m_n}(t_n)\Big) + o_P(1),
\end{aligned}$$

where

$$\mathbf{C}_{1,n}(y) = \int \exp(iyx)\,\mathbf{B}_{1,n}(F(dx)), \quad \mathbf{C}_{2,n}(y) = \int \exp(iyx)\,\mathbf{B}_{2,n}(G(dx)).$$

Hence, it suffices to concentrate on the sequences

$$\begin{aligned}
A_n &= \sqrt{m_n/r_n}\Big(\sqrt{\tau}\,\mathbf{B}_{1,l_n}(\theta\, r_n/m_n) - \theta\,\mathbf{B}_{2,m_n}(r_n/m_n)\Big)\,,\\
B_n &= \begin{pmatrix} \operatorname{Re}\big(z_n\big(\sqrt{\tau}\,\mathbf{C}_{1,l_n}(t_n) - \eta\,\mathbf{C}_{2,m_n}(t_n)\big)\big)\\ \operatorname{Im}\big(z_n\big(\sqrt{\tau}\,\mathbf{C}_{1,l_n}(t_n) - \eta\,\mathbf{C}_{2,m_n}(t_n)\big)\big)\end{pmatrix}.
\end{aligned}$$

By construction of the stochastic integrals $\mathbf{C}_{k,n}(t)$, the vector $(A_n, B_n')'$ is zero mean trivariate Gaussian. In view of (5.18),

$$\operatorname{Var} A_n = \tau\theta(1-\theta\, r_n/m_n) + \theta^2(1-r_n/m_n)\,.$$

The variable $z_n\sqrt{\tau}\,\mathbf{C}_{1,l_n}(t_n) - z_n\eta\,\mathbf{C}_{2,m_n}(t_n)$ is complex Gaussian, having variance $\sigma_n^2 = \tau(1 - |\widetilde{F}(t_n)|^2) + |\eta|^2(1-\widetilde{G}(t_n)|^2)$ and relation $\rho_n = z_n^2\tau(\widetilde{F}(2t_n) - \widetilde{F}(t_n)^2) + z_n^2\eta^2(\widetilde{G}(2t_n) - \widetilde{G}(t_n)^2)\big)$. Hence, both $\operatorname{Var} A_n$ and

$$\operatorname{Cov} B_n = \frac{1}{2}\begin{pmatrix} \sigma_n^2 + \operatorname{Re}\rho_n & \operatorname{Im}\rho_n \\ \operatorname{Im}\rho_n & \sigma_n^2 - \operatorname{Re}\rho_n \end{pmatrix}$$

are uniformly bounded above and uniformly bounded away from zero. For the asymptotic independence of A_n and B_n, it thus remains to show that $\operatorname{Cov}(A_n, B_n) \to 0$, which we will do exemplarily for A_n and B_n^1, the first coordinate of B_n.

First, note that $\mathbf{B}_{1,l_n}$ is independent of $\mathbf{C}_{2,m_n}$, and $\mathbf{B}_{2,m_n}$ is independent of $\mathbf{C}_{1,l_n}$. This yields

$$\begin{aligned}
\operatorname{Cov}(A_n, B_n^1) &= \sqrt{\frac{m_n}{r_n}}\Big(\tau \operatorname{Cov}\big(\mathbf{B}_{1,l_n}(\theta\, r_n/m_n), \operatorname{Re}(z_n \mathbf{C}_{1,l_n}(t_n))\big) \\
&\qquad + \theta \operatorname{Cov}\big(\mathbf{B}_{2,m_n}(r_n/m_n), \operatorname{Re}(\eta\, \mathbf{C}_{2,m_n}(t_n))\big)\Big) \\
&= \sqrt{\frac{m_n}{r_n}}\Big(\tau \operatorname{Re} z_n \operatorname{E}\big(\mathbf{B}_{1,l_n}(\theta\, r_n/m_n)\operatorname{Re}\mathbf{C}_{1,l_n}(t_n)\big) \\
&\qquad - \tau \operatorname{Im} z_n \operatorname{E}\big(\mathbf{B}_{1,l_n}(\theta\, r_n/m_n)\operatorname{Im}\mathbf{C}_{1,l_n}(t_n)\big) \\
&\qquad + \theta \operatorname{Re}\eta \operatorname{E}\big(\mathbf{B}_{2,m_n}(r_n/m_n)\operatorname{Re}\mathbf{C}_{2,m_n}(t_n)\big) \\
&\qquad - \theta \operatorname{Im}\eta \operatorname{E}\big(\mathbf{B}_{2,m_n}(r_n/m_n)\operatorname{Im}\mathbf{C}_{2,m_n}(t_n)\big)\Big).
\end{aligned}$$

Hence, the last four expectations should be $o\big(\sqrt{r_n/m_n}\big)$. Exemplarily again, we only consider the first one. For convenience, set $\mathbf{B} = \mathbf{B}_{1,l_n}$, $\mathbf{C} = \mathbf{C}_{1,l_n}$, and let $\mathbf{W}$ be a standard Brownian motion on $[0,1]$, so that the processes $\mathbf{B}(t)$ and $\mathbf{W}(t) - t\mathbf{W}(1)$ are equal in distribution. With this,

$$\begin{aligned}
\operatorname{E}\big(\mathbf{B}(\theta\, r_n/m_n)\operatorname{Re}\mathbf{C}(t_n)\big) &= \operatorname{E}\Big(\big(\mathbf{W}(\theta\, r_n/m_n) - \frac{r_n}{m_n}\mathbf{W}(1)\big) \\
&\qquad \cdot \Big(\int_0^1 \cos(t_n F^{-1}(y))\,\mathbf{W}(dy) - \mathbf{W}(1)\int_0^1 \cos(t_n F^{-1}(y))\,dy\Big)\Big) \\
&= \int_0^{\theta\, r_n/m_n} \cos(t_n F^{-1}(y))\,dy - \frac{r_n}{m_n}\int_0^1 \cos(t_n F^{-1}(y))\,dy
\end{aligned}$$

which is in fact of the required order. Therefore, A_n and B_n are asymptotically independent, and in view of Lemma 5.3 we have proven Theorem 5.1. □

5.3. Further applications

Testing against tail dominance

The weak limit in (5.17) can also be used to construct a test against tail dominance such as $F(y)/G(y) \to 0$ as $y \to -\infty$. Given $\delta > 0$, consider the hypothesis

$$H_\delta : \theta \le \delta.$$

H_δ will be rejected if θ_n in (5.15) exceeds some level $\kappa_n \geq \delta$. Assume that the rate r_n is chosen sufficiently slow such that $\sqrt{r_n}(\bar{\theta}_n - \theta) \to_p 0$. Then, in view of Theorem 5.1, the type I error can be approximated according to

$$\mathrm{P}_\theta(\theta_n \geq \kappa_n) = \mathrm{P}_\theta\big(\sqrt{r_n}(\theta_n - \theta) \geq \sqrt{r_n}(\kappa_n - \theta)\big) \approx 1 - \Phi\Big(\frac{\sqrt{r_n}(\kappa_n - \theta)}{\sqrt{\tau\theta + \theta^2}}\Big),$$

where Φ denotes the distribution function of the standard normal distribution, so that in order to achieve an approximate level α for all $\theta \leq \delta$, one chooses

$$\kappa_n = \delta + \sqrt{\tau\delta + \delta^2}\, Q_{1-\alpha}/\sqrt{r_n},$$

where $Q_{1-\alpha}$ denotes the standard normal $1-\alpha$–quantile, or one uses the approximate p-value

$$p_n = 1 - \Phi\big(\sqrt{r_n}(\theta_n - \delta)/\sqrt{\tau\delta + \delta^2}\big)$$

in order to reject H_δ if possible.

Distribution functions of regular variation at infinity

Assume that for $b \in \mathbb{R}$ and $a, \alpha, \beta > 0$,

$$F(x) = a|x|^{-\alpha}\big(1 + b|x|^{-\beta} + o(|x|^{-\beta})\big) \qquad \text{as } x \to -\infty.$$

Then F is said to vary regularly at $-\infty$, and α is called the corresponding *exponent of regular variation*. A conditional likelihood estimators for α was already proposed by Hill (1975), and its asymptotic normality was established by Hall (1982). Also, Hall and Welsh (1984) proved that α can not be estimated with a rate faster than $n^{-\beta/(2\beta+\alpha)}$.

Note that $F(x/e)/F(x) \to e^\alpha$ as $x \to -\infty$. Therefore, given an i.i.d. sample $X_1, \ldots, X_n$ drawn from F, Section 5.2, and especially (5.5) and (5.15), now suggest the estimator

$$\alpha_n = \log F_n(h_n/e) - \log G_n(h_n),$$

where F_n and G_n are the empirical distribution functions of $X_1, \ldots, X_{\lfloor \tau n \rfloor}$ and $X_{\lfloor \tau n \rfloor + 1}, \ldots, X_n$, resp., for some $\tau \in (0,1)$, and where h_n is the $\lfloor r_n \rfloor$-th largest order statistic of the second sub-sample.

Corollary 5.1. *If $2\beta/(2\beta+\alpha) > 1/2$, then, for $r_n = n^{\frac{2\beta}{2\beta+\alpha}}$,*

$$\sqrt{r_n}\big(\alpha_n - \alpha\big) \rightsquigarrow \mathcal{N}\big((a(1-\tau))^{-\beta/\alpha} b(e^\beta - 1), 1 + \tau^{-1}(1-\tau)e^{-\alpha}\big).$$

Remark. a. Corollary 5.1 shows that the estimator α_n attains the optimal rate as provided in Hall and Welsh (1984). In fact, from the proof it can be seen that $r_n = n^{2\beta/(2\beta+\alpha)}$ is the fastest rate for which the bias $\sqrt{r_n}(\bar{\theta}_n - \theta)$ corresponding to our estimator α_n does not diverge.

b. The mean squared risk of the weak limit strongly depends on the size of the sub-samples, i. e. the choice of τ. As should be expected, it gets larger the closer τ is chosen to 0 or 1.

It is minimized in τ at the unique solution to the equation

$$\tau^2(1-\tau)^{-\frac{2\beta+\alpha}{\alpha}} = \frac{\alpha a^{2\beta/\alpha}}{2\beta b^2 e^\alpha (e^\beta - 1)^2}\,.$$

Proof of Corollary 5.1. First, since

$$\begin{aligned}\frac{F(h_n/e)}{e^\alpha F(h_n)} - 1 &= \frac{1+b(h_n/e)^{-\beta}+o(|h_n|^{-\beta})}{1+bh_n^{-\beta}+o(|h_n|^{-\beta})} - 1 = \frac{bh_n^{-\beta}(e^\beta-1)+o(|h_n|^{-\beta})}{1+bh_n^{-\beta}+o(|h_n|^{-\beta})} \\ &= \big(b(e^\beta-1)+o_P(1)\big)|h_n|^{-\beta},\end{aligned}$$

for the bias we find that

$$\sqrt{r_n}\Big(\frac{F(h_n/e)}{F(h_n)} - e^\alpha\Big) = \big(be^\alpha(e^\beta-1)+o_P(1)\big)\sqrt{r_n}|h_n|^{-\beta}\,.$$

Set $l_n = \lfloor \tau n \rfloor$ and $m_n = n - l_n$, and let ξ_n denote the (eventually unique) solution giving $F(\xi_n) = \lfloor r_n \rfloor / m_n$. Since $\xi_n \to -\infty$ and $|\xi_n - h_n| = o_P(1)$, see for instance Falk (1989), we particularly have $h_n \sim_p \xi_n$. Further, it is easily seen that $\xi_n \sim -(am_n/r_n)^{1/\alpha}$. In fact, letting $\xi_n' = -(am_n/r_n)^{1/\alpha}$, then

$$\frac{\lfloor r_n \rfloor}{m_n} = F(\xi_n) = a|\xi_n|^{-\alpha}\big(1+b|\xi_n|^{-\beta}+o(|\xi_n|^{-\beta})\big) = \frac{r_n}{m_n}|\xi_n/\xi_n'|^{-\alpha}\big(1+o(1)\big)$$

which implies $\xi_n/\xi_n' \to 1$. Hence, defining $C = a^{-\beta/\alpha} b e^\alpha (e^\beta - 1)$,

$$\begin{aligned}\sqrt{r_n}\Big(\frac{F(h_n/e)}{F(h_n)} - e^\alpha\Big) &= (C+o_P(1))\, n^{\frac{\beta}{2\beta+\alpha}+\frac{2\beta}{2\beta+\alpha}\frac{\beta}{\alpha}} m_n^{-\frac{\beta}{\alpha}} \\ &= (C+o_P(1))\Big(\frac{m_n}{n}\Big)^{-\beta/\alpha} \to_p C(1-\tau)^{-\beta/\alpha}.\end{aligned}$$

Now, using $m_n/l_n \to (1-\tau)/\tau$, Theorem 5.1 handles the variance and provides

$$\sqrt{r_n}\Big(\frac{F_n(h_n/e)}{G_n(h_n)} - e^\alpha\Big) \rightsquigarrow \mathcal{N}\big(C(1-\tau)^{-\beta/\alpha}, \tau^{-1}(1-\tau)e^\alpha + e^{2\alpha}\big).$$

Finally, the function $x \mapsto \log x$ being differentiable at e^α with derivative $e^{-\alpha}$, we get

$$\begin{aligned}\sqrt{r_n}(\alpha_n - \alpha) &= \sqrt{r_n}\Big(\log\frac{F_n(h_n/e)}{G_n(h_n)} - \log e^\alpha\Big) \\ &\rightsquigarrow \mathcal{N}\big(C(1-\tau)^{-\beta/\alpha}e^{-\alpha}, 1+\tau^{-1}(1-\tau)e^{-\alpha}\big)\end{aligned}$$

from the Delta method. □

Bibliography

BELITSER, E. and LEVIT, B. (1995). On minimax filtering over ellipsoids. *Math. Meth. Statist.*, **4** 259–273.

BERAN, R., FEUERVERGER, A. and HALL, P. (1996). On nonparametric estimation of intercept and slope distributions in random coefficient regression. *Ann. Stat.*, **24** 2569–2592.

BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.*, **34** 1204–1232.

BORDES, L. and VANDEKERKHOVE, P. (2010). Semiparametric two-component mixture model with a known component: an asymptotically normal estimator. *Mathematical Methods of Statistics*, **19** 22–41.

BÖTTCHER, A., GRUDSKY, S. M. and MAKSIMENKO, E. A. (2010a). Inside the eigenvalues of certian Hermitian Toeplitz band matrices. *J. Comput. Appl. Math.*, **233** 2245–2264.

BÖTTCHER, A., GRUDSKY, S. M. and MAKSIMENKO, E. A. (2010b). On the structure of the eigenvectors of large hermitian toeplitz band matrices. In *Recent Trends in Toeplitz and Pseudodifferential Operators*, vol. 210 of *Operator Theory: Advances and Applications*. Springer Basel, 15–36.

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Stat.*, **24** 2384–2398.

CAVALIER, L. (1998). Asymptotically efficient estimation in a problem related to tomography. *Math. Methods of Stat.*, **7** 445–456.

CAVALIER, L. and TSYBAKOV, A. B. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Rel.*, **123** 323–354.

CSÖRGŐ, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Probab.*, **9** 130–144.

DAVISON, M. E. (1981). A singular value decomposition for the Radon transform in n-dimensional euclidean space. *Numer. Func. Anal. Opt.*, **3** 321–340.

DAVISON, M. E. (1983). The ill-conditioned nature of the limited angle tomography problem. *Siam J. Appl. Math.*, **43** 428–448.

DAVISON, M. E. and GRUNBAUM, F. A. (1981). Tomographic reconstruction with arbitrary directions. *Communications on Pure and Applied Mathematics*, **34** 77–119.

DEANS, S. R. (1983). *The Radon transform and some of its applications*. John Wiley & Sons.

DEL BARRIO, E., DEHEUVELS, P. and VAN DE GEER, P. (2007). *Lectures on Empirical Processes: Theory and Statistical Applications*. EMS Series of Lectures in Mathematics, European Mathematical Society.

EVERITT, B. S. and HAND, D. J. (1981). *Finite Mixture Distributions*. Chapman and Hall, London.

FAHRMEIR, L., KNEIB, T., LANG, S. and MARX, B. (2013). *Regression: Models, Methods and Applications*. Springer, Berlin, Heidelberg.

FALK, M. (1989). A note on uniform asymptotic normality of intermediate order statistics. *Ann. Inst. Statist. Math.*, **41** 19–29.

FEUERVERGER, A. and HALL, P. (1996). On distribution-free inference for record-value data with trend. *Ann. Stat.*, **24** 2655–2678.

FRIKEL, J. (2013). Sparse regularization in limited angle tomography. *Appl. Comput. Harmon. Anal.*, **34** 117–141.

GAUTIER, E. and KITAMURA, Y. (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica*, **81** 581–607.

GOLUBEV, G. K. and KHASMINSKII, R. Z. (1999). A statistical approach to some inverse problems for partial differential equations. *Probl. Inf. Transm.*, **35** 136–149.

GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz forms and their applications*. Univ. of California Press.

HALL, P. (1981). On the non-parametric estimation of mixture proportions. *J. Roy. Stat. Soc. B Met*, **43** 147–156.

HALL, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc. B*, **44** 37–42.

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.*, **12** 1079–1084.

HALL, P. and ZHOU, X. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.*, **31** 201–224.

HENRY, M., KITAMURA, Y. and SALANIÉ, B. (2010). Identifying finite mixtures in econometric models. *Cowles Foundation Discussion Paper Series*, **1767**.

HILDRETH, C. and HOUCK, J. P. (1968). Some estimators for a linear model with random coefficients. *J. Am. Stat. Assoc.*, **63** 584–595.

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3** 1163–1174.

HODERLEIN, S., KLEMELÄ, J. and MAMMEN, E. (2010). Analyzing the random coefficient model nonparametrically. *Econometric Theory*, **26** 804–837.

HOHMANN, D. and HOLZMANN, H. (2013a). Semiparametric location mixtures with distinct components. *Statistics*, **47** 348–362. Published online January 31, 2012.

HOHMANN, D. and HOLZMANN, H. (2013b). Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation. *Electron. J. Statist.*, **7** 859–880.

HOHMANN, D. and HOLZMANN, H. (2013c). Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation. *Technical Report, Marburg University*. URL `http://www.uni-marburg.de/fb12/stoch/files/holzmann/38.pdf`.

HOHMANN, D. and HOLZMANN, H. (2013d). Weighted angle radon transform: Convergence rates and efficient estimation. *submitted*.

HUNTER, D., WANG, S. and HETTMANSPERGER, T. (2007). Inference for mixtures of symmetric distribution. *Ann. Statist.*, **35** 224–251.

JOHNSTONE, I. M. (1989). On singular value decompositions for the Radon transform and smoothness classes of functions. Tech. Rep. 310, Dept. Statist., Stanford University.

JOHNSTONE, I. M. and SILVERMAN, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Stat.*, **18** 251–280.

KASAHARA, H. and SHIMOTSU, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, **77** 135–175.

KAWATA, T. (1972). *Fourier Analysis in Probability Theory*. Probability and Mathematical Statistics, Academic Press, New York.

KERKYACHARIAN, G., KYRIAZIS, G., LE PENNEC, E., PETRUSHEV, P. and PICARD, D. (2010). Inversion of noisy radon transform by SVD based needlets. *Appl. Comput. Harmon. Anal.*, **28** 24–45.

KERKYACHARIAN, G., LE PENNEC, E. and PICARD, D. (2012). Radon needlet thresholding. *Bernoulli*, **18** 391–433.

KITAMURA, Y. (2004). Nonparametric identifiability of finite mixtures. Unpublished manuscript, Yale University, Department of Economics.

KLEIN, L. R. (1953). *A textbook of econometrics*. Row, Peterson, Evanston.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*, vol. 82 of *Lecture Notes in Statistics*. Springer.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. Wiley series in probability and statistics, John Wiley & Sons, New York.

MURRAY, G. D. and TITTERINGTON, D. M. (1978). Estimation problems with data from a mixture. *J. Roy. Stat. Soc. C App*, **27** 325–334.

NATTERER, F. (1986). *The Mathematics of Computerized Tomography*. John Wiley & Sons.

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and gaussian white noise. *Ann. Stat.*, **24** 2399–2430.

PARTER, S. V. (1961). On the extreme eigenvalues of Toeplitz matrices. *Trans. Amer. Math. Soc.*, **100** 263–276.

PINSKER, M. S. (1980). Optimal filtering of square-integrable signals in Gaussian white noise. *Probl. Peredachi Inf.*, **16** 52–68.

RADON, J. (1917). Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Akad. Wiss.*, **69** 262–277.

RAMANATHAN, T. V. and RAJARSHI, M. B. (1992). Optimal estimation in random coefficient regression models. *Ann. I. Stat. Math.*, **44** 213–227.

REISS, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Stat.*, **36** 1957–1982.

SLEPIAN, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case. *Bell System Technical Journal*, **57** 1371–1430.

SWAMY, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, **38** 311–323.

SWAMY, P. A. V. B. (1971). *Statistical Inference in Random Coefficient Regression Models*. Springer, Berlin.

SZEGÖ, G. (1967). *Orthogonal polynomials*. 3rd ed. Colloquium publications, American Mathematical Society.

TALAGRAND, M. (1996). Lower classes for fractional brownian motion. *J. Theor. Probab.*, **9** 191–213.

TILLI, P. (2003). Universal bounds on the convergence rate of extreme Toeplitz eigenvalues. *Linear Algebra Appl.*, **366** 403–416.

TITTERINGTON, D. M., SMITH, A. F. M. and MAKOV, U. E. (1985). *Statistical analysis of finite mixture distributions*. John Wiley & Sons, Chichester.

VAN DER VAART, A. and WELLNER, J. (2000). *Weak Convergence and Empirical Processes*. Springer Series in Statistics, Springer, New York.

VANDEKERKHOVE, P. (2012). Estimation of a semiparametric mixture of regressions model. *J. Nonparametr. Stat.*, **25** 181–208.

WANG, B. and ZHANG, F. (1992). Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices. *Linear Algebra Appl.*, **160** 113–118.

ZHANG, F. and ZHANG, Q. (2006). Eigenvalue inequalities for matrix product. *IEEE T. Automat. Contr.*, **51** 1506–1509.