Weighted angle Radon transform: Convergence rates and efficient estimation

Daniel Hohmann Hajo Holzmann

January 9, 2015

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

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 $B_1(0) = \{x \in \mathbb{R}^2 : \|x\| \le 1\}$ 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

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

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

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

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 $\gamma > -1/2$,

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

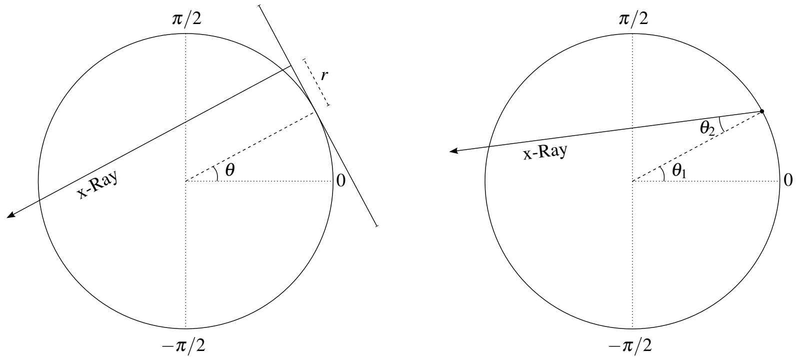
The weight function w_1 in detector space also corresponds to the measurement design, the most important cases being $\gamma = 1$ (fan beam design) as well as $\gamma = 1/2$ (parallel beam design). The weight w_2 with corresponding γ is then required for technical reasons, in order to make the singular 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. Gaussian white noise. This is an idealized statistical model, in which we shall conduct our convergence analysis below. We observe

$$dY(\varphi,s) = \big(\mathrm{R}f\big)(\varphi,s)\,d\mu_1(\varphi,s) + \varepsilon\,dW(\varphi,s), \tag{3}$$



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

Figure 1: 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 means that for any $h(\varphi, s) \in L_2([-\pi/2, \pi/2] \times [-1, 1]; \mu_1)$, we may observe

$$\begin{aligned} Y(h) &= \int_{-\pi/2}^{\pi/2} \int_{-1}^{1} \mathrm{R}f(\varphi, s)\, h(\varphi, s)\, \lambda(\varphi) w_1(s)\, d\varphi\, ds + \varepsilon \int_{-\pi/2}^{\pi/2} \int_{-1}^{1} h(\varphi, s)\, dW(\varphi, s), \\ &= \langle (\mathrm{R}f), h \rangle_{\mu_1} + \varepsilon W(h), \end{aligned} \tag{4}$$

where $W(h)$ is a Gaussian field with mean zero and covariance

$$E\big(W(h_1)W(h_2)\big) = \langle h_1, h_2 \rangle_{\mu_1}, \qquad h_1, h_2 \in L_2([-\pi/2, \pi/2] \times [-1, 1]; \mu_1).$$

For direct observations, regression as well as density estimation problems are asymptotically equivalent to white noise models under fairly general conditions, see Brown and Low (1996), Nussbaum (1996) and Reiß (2008). While no corresponding results are available yet for the indirect models below, the analysis in the technically less complicated white noise model still gives a valuable insight into the difficulty of the estimation problem.

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 $h(\varphi, s)$,

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

which may be estimated unbiasedly from a sample of observations; compare to (4) in the white noise model for analogy.
This model is the 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 1.

For the fan beam design, most statistical literature uses SVD or derived methods (such as needlets), see Cavalier and Tsybakov (2002) or Klemelä and Mammen (2010). In case of parallel beam, Cavalier (1998) uses 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, which arises most naturally in the parallel beam design in form of a limited angle, e.g. in digital breast tomosynthesis, dental tomography or electron microscopy, where the measurement device may only be rotated over a limited range. See Frikel (2013) for further references and also for a discussion of the bias of the filtered back-projection algorithm in case of limited angle.

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 $\mathrm{R}f$. 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.

4. Nonparametric random coefficient regression models. Nonparametric estimation in random coefficient regression models was first studied in Beran, Feuerverger and Hall (1996). These models have recently become quite popular in econometrics, see Hoderlein, Klemelä and Mammen (2010). 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

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

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 $\gamma = 1/2$. Thus,

$$\mathrm{E}\big(h(\Phi,Z)\big) = 2^{-1}\,\langle(\mathrm{R}f_\beta),h\rangle_{\mu_1},$$

which in analogy to the white noise model (4) may be unbiasedly estimated from a sample in this model.

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 2 for an illustration. Our results will

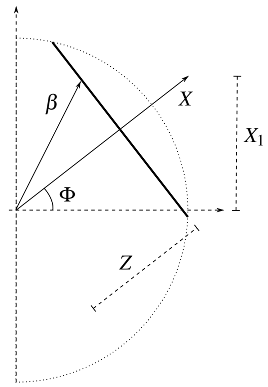


Figure 2: Parametrization in the random coefficient model. The bold line is the set of all $\beta \in B_1(0)$ for which $\beta' X/\|X\| = Z$.

show that for lighter tails, the Radon transform R on the weighted L_2-spaces is in fact more ill-posed.

Main results and structure of the paper

As a first main result, we determine the singular value decomposition of the Radon transform R as an operator between the weighted L_2-spaces in (1), and show how the singular values relate to the eigenvalues of certain Toeplitz matrices associated with the function λ. If we evaluate the white noise model (3) at the singular functions, we obtain a doubly-indexed sequence model. As a next major result, for the limited angle situation, i.e. if $\lambda = \mathbf{1}_{[-\eta,\eta]}$ for an $\eta < \pi/2$, we show that the optimal rate of estimation over ellipsoidal smoothness classes is only logarithmic, and that a simple projection estimator achieves the optimal rate together with the optimal constant. If the weight function has an isolated zero, we give polynomial upper and lower bounds on the rate of convergence, the order of which depends on the degree of the zero. Finally, for functions λ with finite Fourier expansion, we even obtain optimal rates with precise minimax constants in case of the fan-beam design.

The paper is structured as follows. We start Section 2 by reviewing efficient estimation in general white noise sequence models, and in Section 2.2 we introduce the doubly indexed sequence model for the Radon transform. We discuss ellipsoidal smoothness assumptions, and how the Pinsker estimator applies in this model. In Section 2.3 we present the singular values, while the full derivation of the SVD of the weighted angle Radon transform, together with explicit expressions for the singular functions, is given in the supplementary Appendix B.1. Section 3 turns to nonparametric estimation in the sequence model for the Radon transform. We start in Section 3.1 with the severely ill-posed limited angle problem, in which a simple projection estimator is even sharp minimax adaptive. In Section 3.2 we give upper and lower bounds on the rate of convergence in case of polynomial decay of the singular values, and in Section 3.3 we obtain precise rates with asymptotic minimax constants for the fan-beam design ($\gamma = 1$). Section 4 concludes, while proofs are deferred to the supplementary Appendix A. The derivation of the SVD, discussion of the ellipsoidal smoothness assumptions, as well as some further results can be found in the supplementary Appendix B.

2. Gaussian white noise sequence models

2.1. 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). Consider observing

$$Y_k = \theta_k + \varepsilon\sigma_k^{-1}\xi_k, \qquad k = 0, 1, 2, \ldots, \tag{5}$$

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 $\theta = (\theta_0, \theta_1, \ldots)$ 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 $h = (h_0, h_1, \ldots)$, 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 $a = (a_0, a_1, \ldots)$ of real ellipsoid weights, set

$$\Theta = \Theta(a, L) = \Big\{\theta : \sum_{k=0}^{\infty} a_k^2 \theta_k^2 \le L\Big\}. \tag{6}$$

Pinsker estimator. Let $\Theta = \Theta(a, L)$ be an ellipsoid according to (6), and assume that for all $\varepsilon > 0$ there exists a solution c_ε to the equation

$$\varepsilon^2 \sum_{k=0}^{\infty} \sigma_k^{-2} a_k (1 - c_\varepsilon a_k)_+ = c_\varepsilon L, \tag{7}$$

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 $h_k^* = (1 - c_\varepsilon a_k)_+$, $k = 0, 1, \ldots$.

Theorem 2.1 (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*

$$r_\varepsilon^L(\Theta) = \varepsilon^2 \sum_{k=0}^{\infty} \sigma_k^{-2}(1 - c_\varepsilon a_k)_+. \tag{8}$$

b. If

$$\frac{\max_{k: a_k < T} \sigma_k^{-2}}{\sum_{k: a_k < T} \sigma_k^{-2}} \longrightarrow 0 \tag{9}$$

as $T \to \infty$, then $r_\varepsilon(\Theta) \sim r_\varepsilon^L(\Theta)$ as $\varepsilon \to 0$, i. e., under (9) the Pinsker estimator is even asymptotically efficient on $\Theta(a, L)$.

The condition (9) 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 (7) so that the Pinsker estimator is well-defined and Theorem 2.1 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

$$N_\varepsilon = \max\{k : a_k \le c_\varepsilon^{-1}\} = \max\Big\{n : \varepsilon^2 \sum_{k=0}^{n} \sigma_k^{-2} a_k (a_n - a_k) \le L\Big\}, \tag{10}$$

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

$$\theta_k^* = \frac{\varepsilon}{\sigma_k} \sqrt{\frac{(1 - c_\varepsilon a_k)_+}{c_\varepsilon a_k}}. \tag{11}$$

2.2. The sequence model for the Radon transform

Suppose now that we observe the Radon transform $\mathrm{R}f$ of a function $f \in L_2(B_1(0); \mu_2)$ in the white noise model (3).

We require the singular value decomposition of the operator R in (1). It consists of triples

$$\big\{\Psi_{m,l}, \Phi_{m,l}, \sigma_{m,l}\big\}_{m \ge l \ge 0}, \tag{12}$$

where the $(\Psi_{m,l})_{m \ge l \ge 0}$ form an orthonormal basis of $L_2(B_1(0); \mu_2)$, the $\{\Phi_{m,l}\}_{m \ge l \ge 0}$ are orthonormal in $L_2([-\pi/2, \pi/2] \times [-1, 1]; \mu_1)$ and complete in range(R), $\sigma_{m,l} > 0$ for all $m \ge l \ge 0$ and $\mathrm{R}\Psi_{m,l} = \sigma_{m,l}\, \Phi_{m,l}$ and $\mathrm{R}^*\Phi_{m,l} = \sigma_{m,l}\, \Psi_{m,l}$, where R^* is the adjoint operator of R, see Proposition B.2. The $(\Psi_{m,l})_{m \ge l \ge 0}$ and $(\Phi_{m,l})_{m \ge l \ge 0}$ are called the singular functions, the $(\sigma_{m,l})_{m \ge l \ge 0}$ the singular values. The singular values are presented in the next section, while the derivation of the SVD together with explicit forms of the singular functions in terms of orthogonal polynomials, is given in the supplementary Appendix B.1.

Evaluating (4) at the singular functions $\Phi_{m,l}$, we obtain the doubly indexed sequence of observations

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

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 $Y_{m,l} = \sigma_{m,l}^{-1} Y(\Phi_{m,l})$, so that

$$Y_{m,l} = \theta_{m,l} + \varepsilon\sigma_{m,l}^{-1}\xi_{m,l}, \qquad m \geq l \geq 0. \tag{13}$$

Thus, in the doubly indexed sequence model (13), ellipsoidal smoothness assumptions on f correspond to the decay of the Fourier coefficients $\theta_{m,l}$ w.r.t. the basis $(\Psi_{m,l})_{m\geq l\geq 0}$, while rates of convergence depend on the decay of the singular values $\sigma_{m,l}$.

We investigate estimation of θ over the ellipsoids

$$\Theta_1 = \Theta_1(\kappa, L) = \Big\{\theta : \sum_{m\geq l\geq 0} (m+1)^{2\kappa}\theta_{m,l}^2 \leq L\Big\},$$
$$\Theta_2 = \Theta_2(\kappa, L) = \Big\{\theta : \sum_{m\geq l\geq 0} (m-l+1)^{2\kappa}(l+1)^{2\kappa}\theta_{m,l}^2 \leq L\Big\}.$$

Compared to (6), where a is a full sequence of weights, here we use a slightly different notation in which the parameter κ determines the whole weighting sequence.

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

$$\Theta_1(2\kappa, L) \subset \Theta_2(\kappa, L) \subset \Theta_1(\kappa, L). \tag{14}$$

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 also Proposition B.6 for a more general result. A simpler yet natural choice is the ellipsoid Θ_1.

Remark (*Pinsker estimator for the Radon sequence model*). In order to apply Pinkser's Theorem 2.1 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 $l < \tilde l$.

2.3. The singular values

In this subsection we present the singular values $\sigma_{m,l}$ in the SVD (12) of the Radon transform, see Section B in the supplement for the proofs. Let

$$\mathrm{C}_m = \mathrm{diag}(c_{m,0}, \ldots, c_{m,m}), \qquad c_{m,j} = \binom{m}{j}\frac{\Gamma(2\gamma)\Gamma(j+\gamma)\Gamma(m-j+\gamma)}{\Gamma(m+2\gamma)\Gamma(\gamma)^2}, \tag{15}$$

and

$$\mathrm{A}_m = \big(d_{j-k}\big)_{j,k=0,\dots,m}, \qquad m = 0,1,2,\dots, \tag{16}$$

which is the Toeplitz matrix determined by the sequence

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

The Toeplitz matrix A_m is Hermitian and positive semidefinite, and it is well known that it is even positive definite 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. We shall denote its (positive) eigenvalues by

$$\alpha_{m,0} \geq \dots \geq \alpha_{m,m} > 0.$$

The matrix $\mathrm{B}_m := \mathrm{C}_m\mathrm{A}_m$ is then also diagonizable, with strictly positive eigenvalues (see Section B), which we denote by

$$\beta_{m,0} \geq \dots \geq \beta_{m,m} > 0.$$

The singular values of R are given by

$$\sigma_{m,l} = \sqrt{\pi\beta_{m,l}}, \quad m \geq l \geq 0. \tag{17}$$

The case $\gamma = 1$ (Fan beam design). In this case 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

$$\sigma_{m,l} = \sqrt{\frac{\pi\alpha_{m,l}}{m+1}}, \qquad m \geq l \geq 0. \tag{18}$$

Thus, for $\gamma = 1$ the decay of $\sigma_{m,l}$ is determined by the decay of the singular values of the the sequence of Toeplitz matrices A_m 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 present the results that we shall require below in Section 3.

The general case. In the general case, the eigenvalues of B_m cannot be expressed in terms of those of A_m, however, it is possible to derive certain bounds. First, concerning the $c_{m,l}$, and 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 pm for some $p \in (0,1)$, decay according to

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

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 $\gamma \le 1$, the extreme weights satisfy

$$\min_{l=0,\dots,m} c_{m,l} \sim \frac{\Gamma(2\gamma)}{\Gamma(\gamma)^2 4^{\gamma-1}} (m+1)^{-1}, \qquad \max_{l=0,\dots,m} c_{m,l} \sim \frac{\Gamma(2\gamma)}{\Gamma(\gamma)} (m+1)^{-\gamma}. \tag{19}$$

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

From these estimates as well as general bounds on the eigenvalues of products of positive definite Hermitian matrices, see for instance Wang and Zhang (1992) and Zhang and Zhang (2006) we in obtain the bounds

$$\frac{\Gamma(2\gamma)}{\Gamma(\gamma)^2 4^{\gamma-1}} \frac{\alpha_{m,m}}{m+1} (1+o(1)) \overset{(\ge)}{\le} \beta_{m,m} \overset{(\ge)}{\le} \frac{\Gamma(2\gamma)}{\Gamma(\gamma)} \frac{\alpha_{m,m}}{(m+1)^\gamma} (1+o(1)), \qquad -1/2 < \gamma \overset{(\gamma>1)}{\le} 1\,. \tag{20}$$

3. Minimax and efficient estimation for the Radon transform

3.1. Limited angle Radon transform

We start with estimation in the limited angle case, where $\lambda = \mathbf{1}_{[-\eta,\eta]}$ for an $\eta < \pi/2$. In this case the Toeplitz matrices A_m generated by λ are given by

$$\mathrm{A}_m = \Big(\frac{\sin(2(j-k)\eta)}{\pi(j-k)}\Big)_{j,k=0,\dots,m},$$

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

$$\alpha_{m,m} \sim C m^{1/2} e^{-\xi m} \qquad \text{as} \quad m \to \infty, \tag{21}$$

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

$$\xi = \log\Big(1 + \frac{2\sqrt{1-\cos(\pi-2\eta)}}{\sqrt{2} - \sqrt{1-\cos(\pi-2\eta)}}\Big). \tag{22}$$

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

By (20) and (18), this implies exponential decay of $\sigma_{m,m}$ as well, leading to a severely ill-posed inverse problem, see e.g. Mair and Ruymgaart (1996). In this case condition (9) fails to hold and therefore the second part of Pinsker's Theorem 2.1 as stated above does not apply. 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 $0 \le l \le m \le M_\varepsilon$, and $h_{m,l} = 0$ otherwise.

Theorem 3.1. *If there exist $\rho_1, \rho_2 \in \mathbb{R}$ and $\tau_1 \ge \tau_2 > 0$ such that the sequence of smallest singular values $\sigma_{m,m}$ satisfies*

$$m^{\rho_1} e^{-\tau_1 m} \lesssim \sigma_{m,m} \lesssim m^{\rho_2} e^{-\tau_2 m} \qquad \text{as } m \to \infty, \tag{23}$$

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

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

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 (21) as well as the bound given in (20),

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

and we readily arrive at

Corollary 3.2. *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 (22).

Remark. 1. In severely ill-posed problems, the variance is dominated by the bias, even when achieving the optimal constant. Therefore, there are several asymptotically efficient estimators, among them the simple projection estimator.

2. 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.

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 3.3. *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*

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

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 (22).

3.2. Weight functions with isolated zeros

In case of a single root of λ (mod π, typically $\pi/2$), the extreme eigenvalues α_{mm} of the sequence of Toeplitz matrices A_m 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 $\alpha_{m,m}^{-1} \sim C m^\rho$, see Parter (1961). For example, for $\lambda = \cos^2$, $\alpha_{m,m}^{-1} \asymp m^2$. By (20), this implies polynomial decay of the singular values $\sigma_{m,l}$ as well.

First we state the following general result.

Proposition 3.4. *a. If there exists $\rho \ge 0$ 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 $0 \le \rho_1 \le \rho < \rho_1 + 1$. If

$$m^{-\rho} \lesssim \beta_{m,m} \lesssim m^{-\rho_1} \qquad \text{as } m \to \infty, \tag{24}$$

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

$$\beta_{m,m}^{-1} \sim C m^\rho \qquad \text{as } m \to \infty, \tag{25}$$

then

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

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. If the minimal eigenvalue $\alpha_{m,m} \asymp m^{-\tilde{\rho}}$, then from the estimate in (20), the condition of a. is satisfied with $\rho = \tilde{\rho} + 1$ in case $-1/2 < \gamma \le 1$, as well as $\rho = \tilde{\rho} + \gamma$ for $\gamma > 1$. Further, (24) 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 (25) we require $\gamma = 1$.

3.3. 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 $\beta_{m,m}$ 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.

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 (18). We shall impose the following assumptions on the eigenvalues $\alpha_{m,l}$ of the Toeplitz matrices A_m.

Assumption A1. There exist $C > 0$ and $\rho \geq 1$ such that

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

Assumption A2. 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 corresponds to that of Section 3.2.

First we show that the above conditions are satisfied in certain specific cases. We say that λ is banded if

$$\lambda(\varphi) = \sum_{k=-r}^{r} d_k e^{i2k\varphi}, \qquad r \in \mathbb{N}, \quad d_r \neq 0, \quad \bar{d_k} = d_{-k},$$

since, by construction, the Hermitian Toeplitz matrices A_m generated by λ are banded in this case, and in fact, the coefficients d_k are exactly the entries of A_m. In particular, the condition $\bar{d_k} = d_{-k}$ ensures that λ is real.

Using the results of Böttcher, Grudsky and Maksimenko (2010a) on the uniform behavior of the eigenvalues of banded Toeplitz matrices A_m, we obtain

Proposition 3.5. *Suppose that λ is banded and satisfies $\lambda(-\pi/2) = \lambda(\pi/2) = 0$. Further, assume that 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. Then the eigenvalues $\alpha_{m,l}$ satisfy Assumption A1 with $\rho = 3$ and $C = 4/(3\lambda''(\pi/2))$, as well as Assumption A2 with $\rho = 3$ and $c_j = \frac{8}{\lambda''(\pi/2)\pi^2}(j+1)^{-2}$.*

Linear Minimax risk on Θ_1 under A1. Let $a_{m,l} = (m+1)^{\kappa}$ be the ellipsoid weights corresponding to $\Theta_1(\kappa, L)$. From (10) 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 ot Section 2. 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

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

By A1 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 (10), and minding that $(1 - c_\varepsilon a_{m,l})_+ = 0$ for $m > N_\varepsilon$, from Pinsker's theorem we obtain

$$\begin{aligned} r_\varepsilon^L(\Theta_1(\kappa,L)) &\sim \varepsilon^2 \sum_{m=0}^{N_\varepsilon}\sum_{l=0}^{m} \sigma_{m,l}^{-2}(1 - N_\varepsilon^{-\kappa}(m+1)^\kappa)\\ &\sim \frac{C\varepsilon^2}{\pi}\sum_{m=0}^{N_\varepsilon}\big(m^\rho - N_\varepsilon^{-\kappa} m^{\kappa+\rho}\big)\\ &\sim \frac{C\varepsilon^2}{\pi}\frac{\kappa}{(\rho+1)(\kappa+\rho+1)} N_\varepsilon^{\rho+1}\\ &\sim C_1^*\,\varepsilon^{\frac{4\kappa}{2\kappa+\rho+1}}\end{aligned} \tag{26}$$

with $C_1^* = C_1^*(\kappa,\rho,L,C)$ given in Theorem 3.6 below.

Linear Minimax risk on Θ_2 under A2. 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 A2 then reads

$$\alpha_{j+k,k}^{-1} = c_j k^{\rho-1} + O\big(((j+1)(k+1))^{\rho-1-\delta}\big), \qquad j,k \ge 0. \tag{27}$$

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 (10) we have $\big\{(j,k) \prec_2 (j,k)_\varepsilon\big\} \cup \{(j,k)_\varepsilon\} =$

(N_ε), 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 A.6 in Section A.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

$$K(\rho,c) = \sum_{j=0}^{\infty} c_j (j+1)^{-(\rho+1)}. \tag{28}$$

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 (26) and using Lemma A.6, we find that

$$\begin{aligned} r_\varepsilon^L(\Theta_2(\kappa,L)) &= \varepsilon^2 \sum_{(j,k)\in(N_\varepsilon)} \sigma_{j+k,k}^{-2}(1 - N_\varepsilon^{-\kappa} a_{j+k,k}) \\ &\sim C_2^* \,\varepsilon^{\frac{4\kappa}{2\kappa+\rho+1}} \end{aligned} \tag{29}$$

with $C_2^* = C_2^*(\kappa,\rho,L,c)$ given in Theorem 3.6 below.

Asymptotic efficiency on Θ_1 *and* Θ_2. Given (26) and (29), we now easily arrive at

Theorem 3.6. *For $i = 1,2$, under A1 and A2, respectively,*

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

where

$$\begin{aligned} C_i^* &= \Big(\frac{\Xi_i \kappa}{\pi(\kappa+\rho+1)}\Big)^{\frac{2\kappa}{2\kappa+\rho+1}} \frac{\big(L(2\kappa+\rho+1)\big)^{\frac{\rho+1}{2\kappa+\rho+1}}}{\rho+1}, \\ \Xi_i &= \begin{cases} C, & i=1, \\ K(\rho,c), & i=2. \end{cases} \end{aligned} \tag{30}$$

Example 3.7. For the ordinary Radon transform, i. e. $\lambda = 1$, we have $\sum_{l=0}^m \alpha_{m,l}^{-1} = m+1$, whence A1 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. For a banded weight function λ satisfying the assumptions of Proposition 3.5, Theorem 3.6 implies 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. Concluding remarks

- 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.
- 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 B.3 in the supplement, where we also derive the asymptotic minimax constants.
- In order to avoid computation of the spectral data, iterative methods might be an alternative, see Bissantz et al. (2007).
- The case of an unknown weight function, for which additional data is available (e.g. in the random coefficients model) leads to a problem with noisy operator, as studied in Hoffmann and Reiss (2008).
- In higher dimensions, injectivity of the limited angle Radon transform as well as the analytic form of its SVD seem not be established.

Acknowledgements

The authors would like to thank the referee and the editor for helpful comments, as well as Thorsten Hohage for helpful discussions.

References

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.

BISSANTZ, N., HOHAGE, T., MUNK, A. and RUYMGAART, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, **45** 2610–2636.

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.

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.

FRIKEL, J. (2013). Sparse regularization in limited angle tomography. *Appl. Comput. Harmon. Anal.*, **34** 117–141.

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.

HODERLEIN, S., KLEMELÄ, J. and MAMMEN, E. (2010). Analyzing the random coefficient model nonparametrically. *Econometric Theory*, **26** 804–837.

HOFFMANN, M. and REISS, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Stat.*, **36** 310–336.

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.

KLEMELÄ, J. and MAMMEN, E. (2010). Empirical risk minimization in inverse problems. *Ann. Stat.*, **38** 482–511.

MAIR, B. A. and RUYMGAART, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, **56** 1424–1444.

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.

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.

SZEGÖ, G. (1967). *Orthogonal polynomials*. 3rd ed. Colloquium publications, American Mathematical Society.

TILLI, P. (2003). Universal bounds on the convergence rate of extreme Toeplitz eigenvalues. *Linear Algebra Appl.*, **366** 403–416.

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.

A. Proofs

A.1. Proofs for Section 3.1

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 A.2 below). The main ingredient is the following lemma.

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

$$\mathrm{E}\big[4(1+\exp(2Z))^{-2}\big] \geq 1-2t. \tag{31}$$

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 (31) and thus concludes the proof. □

Lemma A.2. *For any ellipsoid Θ, the minimax risk in sequence model* (5) *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

$$r_\varepsilon(\Theta) \geq \sum_{k=0}^{\infty} \inf_{\hat{\theta}_k} \int_\Theta \mathrm{E}_\theta\left[(\hat{\theta}_k - \theta_k)^2\right] \pi(d\theta). \tag{32}$$

Now for any $X = (X_k)_{k\geq 0} \sim \pi$ such that $(Y_k, X_k)_{k\geq 0}$ are independent and such that $Y_k|X_k \sim \mathcal{N}(X_k, \varepsilon^2\sigma_k^{-2})$, by sufficiency, the Bayes risks in (32) are minimized by $\hat{\theta}_k = \mathrm{E}[X_k|Y_k]$, so that the conclusion follows from Lemma A.1. $\square$

Lemma A.3. *Consider the sequence model* (5) *and the ellipsoid* $\Theta(a,L)$ *according to* (6) *with* $a_k = (k+1)^\kappa$*. If there exist* $\gamma_1, \gamma_2 > 1$ *such that*

$$\liminf_{k\to\infty} \sigma_k/\sigma_{k+1} \geq \gamma_1, \qquad \limsup_{k\to\infty} \sigma_k/\sigma_{k+1} \leq \gamma_2, \tag{33}$$

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* (11).

Proof. First, we may rewrite

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

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 (33) 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 $1 - (1-x)^\kappa \leq \max(1,\kappa)x$, $0 \leq x \leq 1$, as well as

$c_\varepsilon > a_{N_\varepsilon+1}^{-1}$ and (33) again, the second sum satisfies

$$
\begin{aligned}
S_{2,\varepsilon} &\lesssim \sum_{k=n_\varepsilon+1}^{N_\varepsilon} \sigma_k^{-2}(1-c_\varepsilon a_k)^2 = \sum_{j=0}^{N_\varepsilon-n_\varepsilon-1} \sigma_{N_\varepsilon-j}^{-2}(1-c_\varepsilon a_{N_\varepsilon-j})^2 \\
&\lesssim \sigma_{N_\varepsilon}^{-2} \sum_{j=0}^{N_\varepsilon-n_\varepsilon-1} \gamma_1^{-2j}\Big(1-\Big(\frac{N_\varepsilon-j+1}{N_\varepsilon+2}\Big)^\kappa\Big)^2 \lesssim \sigma_{N_\varepsilon}^{-2} \sum_{j=0}^{N_\varepsilon-n_\varepsilon-1} \gamma_1^{-2j}\Big(\frac{j+1}{N_\varepsilon+2}\Big)^2 \\
&\lesssim \sigma_{N_\varepsilon}^{-2} N_\varepsilon^{-2}.
\end{aligned} \tag{34}
$$

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 (33),

$$
\begin{aligned}
r_\varepsilon^L(\Theta) &= \varepsilon^2 \sum_{j=0}^{N_\varepsilon} \sigma_{N_\varepsilon-j}^{-2}(1-c_\varepsilon a_{N_\varepsilon-j}) \gtrsim \varepsilon^2 \sigma_{N_\varepsilon}^{-2} \sum_{j=0}^{N_\varepsilon} \gamma_2^{-2j}\Big(1-\Big(\frac{N_\varepsilon+1-j}{N_\varepsilon+1}\Big)^\kappa\Big) \\
&\gtrsim \varepsilon^2 \sigma_{N_\varepsilon}^{-2}(N_\varepsilon+1)^{-1} \sum_{j=0}^{N_\varepsilon} \gamma_2^{-2j} j \\
&\gtrsim \varepsilon^2 \sigma_{N_\varepsilon}^{-2} N_\varepsilon^{-1}.
\end{aligned} \tag{35}
$$

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 3.1. 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

$$
\tilde\Theta = \tilde\Theta(\kappa, L) = \Big\{\theta : \sum_{m=0}^{\infty} (m+1)^{2\kappa} \theta_{m,m}^2 \le L, \theta_{m,l} = 0, m \ne l\Big\}, \tag{36}
$$

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 A.2 and (23), 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 (33) is satisfied. In fact, letting $\tilde{\theta}^*$ be the Pinsker solution according to (11) 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 A.3 it follows that

$$\varepsilon^{-2} \sum_{m=0}^{\infty} (\tilde{\theta}_{m,m}^*)^4 \tilde{\sigma}_{m,m}^2 = o(\tilde{r}_\varepsilon^L(\tilde{\Theta})), \tag{37}$$

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 (7) and (10) 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 (37), 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 A.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

$$
\begin{aligned}
\sup_{\theta\in\Theta_i} \sum_{m=M_\varepsilon+1}^{\infty} \sum_{l=0}^{m} \theta_{m,l}^2 \le \sup_{\theta\in\Theta_i} M_\varepsilon^{-2\kappa} \sum_{m=M_\varepsilon+1}^{\infty} \sum_{l=0}^{m} (m+1)^{2\kappa} \theta_{m,l}^2 \le L M_\varepsilon^{-2\kappa}, \\
\sum_{m=0}^{n} \sum_{l=0}^{m} \sigma_{m,l}^{-2} \le \sum_{m=0}^{n} (m+1) \sigma_{m,m}^{-2} \lesssim \sum_{m=0}^{n} m^{1-2\rho_1} e^{2\tau_1 m} \lesssim n^{1-2\rho_1} e^{2\tau_1 n},
\end{aligned}
\tag{38}
$$

where we used Lemma A.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 A.4. *For all $\gamma > 1$ and $\delta, c_1, c_2 \in \mathbb{R}$,*

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

Proof of Lemma A.4. 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,

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

so that

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

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

Proof of Corollary 3.3. 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 (22). Using the bound given in (20) as well as the first part of Theorem 3.1 finishes the proof. □

A.2. Proofs for Section 3.2

Proof of Proposition 3.4. a. Because of the inclusion relation (14), it suffices to consider $i = 1$. As in Theorem 3.1, consider a projection estimator $\hat{\theta}(h^{Pr})$ with truncation level M_ε. Its bias is estimated in (38), while the variance term may be bounded by

$$\sum_{m=0}^{M_\varepsilon} \sum_{l=0}^{m} \sigma_{m,l}^{-2} \leq \sum_{m=0}^{M_\varepsilon} (m+1)\sigma_{m,m}^{-2} \lesssim \sum_{m=0}^{M_\varepsilon} m^{\rho+1} \lesssim M_\varepsilon^{\rho+2}, \tag{39}$$

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

$$\frac{\sigma_{n,n}^{-2}}{\sum_{m=0}^{n}\sum_{l=0}^{m}\sigma_{m,l}^{-2}} \le \frac{\sigma_{n,n}^{-2}}{\sum_{m=0}^{n}\sigma_{m,m}^{-2}} = O(n^{\rho-\rho_1-1}) = o(1), \quad n\to\infty, \tag{40}$$

condition (9) is satified, 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 3.1, we at once obtain the lower bound

$$r_\varepsilon(\Theta_i) \ge r_\varepsilon(\tilde\Theta)$$

with reduced ellipsoid $\tilde\Theta=\tilde\Theta(\kappa,L)$ defined in (36).

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 $(\theta_{0,0},\theta_{1,1},\ldots)$ within the ellipsoid $\Theta(a,L)$ defined in (6) 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

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

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 (24) 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, (40) shows that condition (9) 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 (25) we find the exact rates

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

and

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

□

A.3. Proofs for Section 3.3

Under the assumptions of Proposition 3.5, 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

$$\alpha_{m,l} \geq C_\varepsilon \tag{41}$$

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

$$\alpha_{m,l} = \frac{\lambda''(\pi/2)\pi^2}{8}\Big(\frac{m-l+1}{m+2}\Big)^2 + O\Big(\Big(\frac{m-l+1}{m+2}\Big)^3\Big) \tag{42}$$

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

Lemma A.5. *If λ is banded and satisfies the assumptions of Proposition 3.5 holds, then there exists a constant $C > 0$ such that the eigenvalues $\alpha_{m,l}$ of the Toeplitz matrices A_m generated by g satisfy*

$$\Big|\alpha_{m,l}^{-1} - \frac{8}{\lambda''(\pi/2)\pi^2}\Big(\frac{m+2}{m-l+1}\Big)^2\Big| \leq C\frac{m+2}{m-l+1}, \qquad m \geq l \geq 0.$$

Proof. Set $c = 8/(\lambda''(\pi/2)\pi^2)$ and $\Delta_{m,l} = \big|\alpha_{m,l}^{-1} - c\big(\frac{m+2}{m-l+1}\big)^2\big|$. For the small eigenvalues $\alpha_{m,l}$, (42) 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 (41), 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. $\square$

Remark. In order to obtain (41) and (42) 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 3.5. In order to show the statement concerning Assumption A1, in view of Lemma A.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 A2, from Lemma A.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 A2 holds true for any $\delta \leq 1$. $\square$

Lemma A.6. *If there exist $\beta \geq 1$, $\delta > 0$, and a positive, bounded sequence $c = (c_0, c_1, \ldots)$ 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 $K(\beta, c) = \sum_{j=0}^{\infty} c_j (j+1)^{-(\beta+1)}$.

Proof of Lemma A.6. Conveniently, assume that $\delta \leq 1$, and set $[n] = \{(j,k) : j,k \geq 1, jk \leq n\}$, $\bar{\alpha}_{j+k,k} = \alpha_{j+k-2,k-1}$, 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, (**??**) gives $\bar{\alpha}_{j+k,k} = \bar{c}_j k^{\beta} + O\big(((j+1)(k+1))^{\beta-\delta}\big)$, so for any $x \ge 1, \gamma \ge 0$, 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. □

Proof of Theorem 3.6. In view of (26) and (29), it remains to show that condition (9) holds. Under A1,

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

and

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

And under A2,

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

while Lemma A.6 shows that

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

So, evidently, in both cases (9) holds. □

B. Appendix

B.1. The 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 (1).

We start with the following two results.

Proposition B.1. *If λ is integrable, the Radon transform* R *as a map between the weighted L_2-spaces in (1) is continuous with operator 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

$$\mathrm{R}_{\varphi} : L_2(B_1(0);\mu_2) \longrightarrow L_2([-1,1]; w_1(s)\,ds) \tag{43}$$

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

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

□

Proposition B.2. *The adjoint operator of* R *is given by*

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

Proof. For $\varphi \in [-\pi/2,\pi/2]$ fixed, let the operator R_φ, as in (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

$$(\mathrm{R}^*g)(x,y) = \int_{-\pi/2}^{\pi/2} (\mathrm{R}_\varphi^* g_\varphi)(x,y)\,\lambda(\varphi)\,d\varphi, \tag{44}$$

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. □

Next let us introduce the ingredients of the singular value decomposition. For the Toeplitz matrix A_m in (16), let

$$\{\mathrm{v}_{m,l} = (v_{m,l}^{(0)},\ldots,v_{m,l}^{(m)})'\}_{l=0}^m$$

denote an orthonormal basis of eigenvectors corresponding to the real eigenvalues $\alpha_{m,0} \geq \ldots \geq \alpha_{m,m} > 0$. Define the matrices

$$\mathrm{V}_m = \big(\mathrm{v}_{m,0},\ldots,\mathrm{v}_{m,m}\big), \qquad \Lambda_m = \mathrm{diag}(\alpha_{m,0},\ldots,\alpha_{m,m}),$$

Let C_m be defined in (15), and let

$$\mathrm{B}_m = \Lambda_m^{1/2} \mathrm{V}_m^* \mathrm{C}_m \mathrm{V}_m \Lambda_m^{1/2}, \tag{45}$$

a Hermitian matrix which is similar to $\mathrm{C}_m \mathrm{A}_m$, and hence has the same eigenvalues. Let

$$\{\mathrm{w}_{m,l} = (w_{m,l}^{(0)}, \dots, w_{m,l}^{(m)})'\}_{l=0}^m$$

denote an orthonormal basis of eigenvectors of B_m corresponding to the eigenvalues $\beta_{m,0}, \dots, \beta_{m,m} > 0$.

For $m \geq l \geq 0$, let $h_{m,l}(\varphi) = e^{-i(m-2l)\varphi}$, and let

$$\tilde{\tilde{h}}_{m,l} = \mathrm{w}_{m,l}' \tilde{\mathrm{h}}_m = \sum_{k_1,k_2=0}^{m} \frac{w_{m,l}^{(k_1)} v_{m,k_1}^{(k_2)}}{\sqrt{\pi \alpha_{m,k_1}}} h_{m,k_2}. \tag{46}$$

Let C_m^γ denote the Gegenbauer or ultraspherical polynomials on $[-1,1]$, and let

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

where w_1 is defined in (2). The ϕ_m are orthogonal and complete in $L_2([-1,1]; w_1(s)\,ds)$, with

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

see Davison (1983). For $m \geq l \geq 0$, let

$$\Phi_{m,l}(\varphi, s) = \frac{\phi_m(s)}{\|\phi_m\|_{w_1}} \tilde{\tilde{h}}_{m,l}(\varphi), \qquad -\pi/2 \leq \varphi \leq \pi/2, \, -1 \leq s \leq 1, \tag{47}$$

Further, let $P_n^{(\alpha,\beta)}$ denote the Jacobi polynomials, and for $(x,y) = r e^{i\theta}$ let

$$\widetilde{\Psi}_{m,l}(x,y) = \frac{h_{m,l}(\theta) J_{m,l}(r)}{w_2(x,y)}, \qquad J_{m,l}(r) = \frac{\pi \Gamma\big(\gamma+m-l\big)}{(m-l)!\,\Gamma(\gamma)} r^{m-2l} P_l^{(\gamma-1,m-2l)}(2r^2-1), \tag{48}$$

and

$$\Psi_{m,l}(x,y) = \frac{\sqrt{\beta_{m,l}}}{\pi\sqrt{d_m}} \sum_{k_1,k_2=0}^{m} \frac{w_{m,l}^{(k_1)} v_{m,k_1}^{(k_2)}}{c_{m,k_2}\sqrt{\alpha_{m,k_1}}} \widetilde{\Psi}_{m,k_2}(x,y), \tag{49}$$

Theorem B.3. *Set $\sigma_{m,l} = \sqrt{\pi \beta_{m,l}}$ as in (17). The singular value decomposition of* R *between the weighted L_2-spaces in (1) is given by*

$$\big\{\Psi_{m,l}, \Phi_{m,l}, \sigma_{m,l}\big\}_{m \geq l \geq 0},$$

where $\Phi_{m,l}$ is defined in (47), and $\Psi_{m,l}$ is defined in (49). 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*

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

Proof of Theorem B.3. We start with the following lemma.

Lemma B.4. *For $\phi_m = w_1^{-1} C_m^\gamma$ and $h \in L_2([-\pi/2,\pi/2];\lambda(\varphi)d\varphi)$, 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 (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 B.4 constitutes the first step to determine the spectral decomposition of the operator RR^* and hence the SVD of R. It shows that RR^* leaves the subspaces V_m of $L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1)$ with

$$V_m = \left\{h(\varphi)\,\phi_m(s),\ h \in L_2([-\pi/2,\pi/2];\lambda(\varphi)d\varphi)\right\}$$

invariant. Therefore, in the next lemma we study the action of the self-adjoint integral operators T_m on $L_2([-\pi/2,\pi/2],\lambda(\varphi)d\varphi)$ given by

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

Lemma B.5. *Then the following statements hold:*

1. *T_m vanishes on the orthogonal complement of $\mathrm{lin}\{h_{m,l}\}_{l=0}^m$, and $T_m \mathrm{h}_m = \pi(\mathrm{C}_m\mathrm{A}_m)'\mathrm{h}_m$.*
2. *The functions*

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

are an orthonormal basis of

$$\mathrm{lin}\{h_{m,l}\}_{l=0}^m \subset L_2([-\pi/2,\pi/2];\lambda(\varphi)d\varphi),$$

and

$$T_m \tilde{\mathrm{h}}_m = \pi \mathrm{B}'_m \tilde{\mathrm{h}}_m, \quad \textit{where } \tilde{\mathrm{h}}_m = (\tilde{h}_{m,0}, \ldots, \tilde{h}_{m,m})', \quad T_m \tilde{\mathrm{h}}_m = (T_m \tilde{h}_{m,0}, \ldots, T_m \tilde{h}_{m,m})',$$

and B_m *is defined in (45).*

3. *The functions* $\big(\tilde{\tilde{h}}_{m,l}\big)_{l=0}^m$*, defined in (46), form an orthonormal basis of* $\operatorname{lin}\{h_{m,l}\}_{l=0}^m$ *as well and*
$$T_m \tilde{\tilde{h}}_{m,l} = \pi \beta_{m,l} \tilde{\tilde{h}}_{m,l}.$$

Proof. Ad 1.: In view of (4.9.19) and (4.9.21) in Szegö (1967), the polynomials $C_m^\gamma(\cos\varphi)$ attain the explicit form

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

so that, since $C_m^\gamma(1) = \Gamma(m+2\gamma)/(\Gamma(2\gamma)m!)$, 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 $\operatorname{lin}\{h_{m,0}, \ldots, h_{m,m}\}$ in $L_2([-\pi/2, \pi/2]; \lambda(\varphi)d\varphi)$, and defining

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

we find that

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

proving part 1.

Ad 2.: Orthonormality of the functions $\tilde{h}_{m,0} \ldots, \tilde{h}_{m,m}$ follows from that of $\mathrm{v}_{m,0}, \ldots, \mathrm{v}_{m,m}$. 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

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

This in particular implies that $\tilde{h}_{m,0},\dots,\tilde{h}_{m,m}$ are linearly independent so that, since $\tilde{h}_{m,l} \in \mathrm{lin}\{h_{m,l}\}_{l=0}^{m}$, $l=0,\dots,m$, 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 2.

Part 3. is proved similarly as part 2. $\square$

From Lemma B.4 and Lemma B.5, part 3., it follows that for all $m \geq l \geq 0$,

$$\mathrm{R}\mathrm{R}^*\Phi_{m,l} = \tilde{\phi}_m T_m \tilde{\tilde{h}}_{m,l} = \pi\beta_{m,l}\Phi_{m,l}. \tag{50}$$

Further, from Lemma B.5, parts 1. and 3., the system $\{\Phi_{m,l}\}_{m\geq l\geq 0}$ is orthonormal and complete in the orthogonal complement of the kernel of R^*, and hence in the closure of $\mathrm{range}(\mathrm{R})$.

Setting $\sigma_{m,l} = \sqrt{\pi\beta_{m,l}}$ and $\Psi_{m,l} = \sigma_{m,l}^{-1}\mathrm{R}^*\Phi_{m,l}$, it follows from (50) that

$$\mathrm{R}\Psi_{m,l} = \sigma_{m,l}\,\Phi_{m,l}, \qquad \mathrm{R}^*\Phi_{m,l} = \sigma_{m,l}\,\Psi_{m,l}.$$

To complete the proof of the theorem, it remains to show (49) and that the $\{\Psi_{m,l}\}_{0\leq l\leq m}$ form an orthonormal basis of $L_2(B_1(0);\mu_2)$.

The functions $(\widetilde{\Psi}_{m,l})_{0\leq l\leq m}$ in (48) form an orthogonal basis of $L_2(B_1(0);\mu_2)$. Call the functions on the right side of (49) $\widehat{\Psi}_{m,l}(x,y)$. By orthonormality of the vectors $v_{m,l}$ and $w_{m,l}$, it follows that the $(\widehat{\Psi}_{m,l})_{0\leq l\leq m}$ form an orthogonal basis of $L_2(B_1(0);\mu_2)$ as well.

From Davison (1983, theorem 3.2),

$$\big(\mathrm{R}\widetilde{\Psi}_{m,l}\big)(\varphi,s) = \pi c_{m,l}\,h_{m,l}(\varphi)\,\phi_m(s). \tag{51}$$

Further by (51) and the definitions of $\Phi_{m,l}$ and $\hat{\Psi}_{m,l}$, we have that $\big(\mathrm{R}\hat{\Psi}_{m,l}\big) = \sigma_{m,l}\Phi_{m,l}$. Since the $(\Phi_{m,l})_{0\leq l\leq m}$ are orthonormal in $L_2([-\pi/2,\pi/2]\times[-1,1];\mu_1)$, this implies that R as an operator between the weighted L_2 spaces in (1) is injective. By (50), 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. This concludes the proof of Theorem B.3. $\square$

B.2. Singular functions and smoothness conditions in case $\gamma = 1$

First we specialize our results for the singular functions to 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 $\sigma_{m,l} = \sqrt{\pi\alpha_{m,l}/(m+1)}$, $m \geq l \geq 0$. 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 $\{\mathrm{v}_{m,l} = (v_{m,l}^{(0)},\ldots,v_{m,l}^{(m)})'\}_{l=0}^{m}$ 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 (49) are then expressed as

$$\Psi_{m,l}(x,y) = \frac{\sqrt{m+1}}{\pi}\sum_{k=0}^{m} v_{m,l}^{(k)} z_{m,k}(x,y), \qquad m \geq l \geq 0. \tag{52}$$

Next, 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 B.6. *In case $\gamma = 1$, 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* (52)*, satisfy*

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

Proposition B.6 motivates us to consider

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

where $\mathrm{v}_{m,l} = (v_{m,l}^{(0)},\ldots,v_{m,l}^{(m)})'$ are the orthonormal eigenvectors of A_m. 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 therefore focused on the smoothness classes Θ_1 and Θ_2, but point out the inclusion relations

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

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

Proof of Proposition B.6. In order to deduce the summability condition of Proposition B.6, similar as in Johnstone (1989) we differentiate the singular functions $\Psi_{m,l}$ given in (52) 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

$$D^p \bar{D}^q z_{m,l} = \begin{cases} \frac{s!}{\pi} h_{m-s,l-p}^{s+1}, & m-q \ge l \ge p, \\ 0, & \text{else}, \end{cases} \tag{53}$$

where

$$h_{m,l}^{\gamma}(x,y) = \int_{-\pi/2}^{\pi/2} C_m^{\gamma}(x\cos\varphi + y\sin\varphi) e^{-i(m-2l)\varphi} d\varphi, \tag{54}$$

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

$$\|D^p \bar{D}^q z_{m,l}\|_{\mu_3}^2 = \frac{\pi^{1/2}(s+1)(2s+1)!}{2^{2s+1} s! \Gamma(s+3/2)} \frac{(m-l+p)!(l+q)!}{(l-p)!(m-l-q)!(m+1)}. \tag{55}$$

Now, it suffices to show that the summability condition

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

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,

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

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

$$\begin{aligned} \frac{(m-l+p)!}{(m-l-q)!} &\ge (m-l-q+1)^s \ge (m-l+1)^s (s+1)^{-s}, \\ \frac{(l+q)!}{(l-p)!} &\ge (l-p+1)^s \ge (l+1)^s (s+1)^{-s}. \end{aligned}$$

Therefore, by (55) there exist constants $c_s, C_s > 0$, only depending on $s = p+q$, such that

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

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

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 (56) below, the weak derivatives of f with respect to the operators D and $\bar{D}$ satisfy

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

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 (55). 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

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

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 B.2, the eigenfunctions in brain space can be written as

$$\begin{aligned}\Psi^\gamma_{m,l}(x,y) &= \frac{1}{\pi\sqrt{d_m^\gamma c^\gamma_{m,l}}\,w_2^\gamma(x,y)}\int_{-\pi/2}^{\pi/2}C_m^\gamma(x\cos\varphi+y\sin\varphi)e^{-i(m-2l)\varphi}\,d\varphi\\ &= \frac{h^\gamma_{m,l}(x,y)}{\pi\sqrt{d_m^\gamma c^\gamma_{m,l}}\,w_2^\gamma(x,y)}\end{aligned}$$

with $h^\gamma_{m,l}$ defined in (54), and in particular, regarding (52) and minding that $d_m^1=1$, $c^1_{m,l}=(m+1)^{-1}$, and $w_2^1(x,y)=\pi$, the Zernike functions are given by

$$z_{m,l}(x,y) = \frac{\pi}{\sqrt{m+1}}\Psi^1_{m,l}(x,y) = \pi^{-1}h^1_{m,l}(x,y). \tag{56}$$

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 $d/dsC_m^\gamma(s) = 2\gamma C_{m-1}^{\gamma+1}(s)$, see e. g. (4.7.14) in Szegö (1967), it readily follows that

$$Dh_{m,l}^\gamma = \gamma h_{m-1,l-1}^{\gamma+1}, \qquad \bar{D}h_{m,l}^\gamma = \gamma h_{m-1,l}^{\gamma+1},$$

where in particular $Dh_{m,0}^\gamma = \bar{D}h_{m,m}^\gamma = 0$. For $p,q \in \mathbb{N}$ such that $p+q=s$ this provides (53).

The norm of these derivatives can now be evaluated with respect to μ_3^s. For this, note that $w_3^\gamma = (w_2^{\gamma+1})^{-1}$ and that the $\Psi_{m,l}^\gamma$ are normalized with respect to μ_2^γ. Therefore,

$$\begin{aligned}\|h_{m,l}^{\gamma+1}\|_{\mu_3^\gamma} &= \pi\sqrt{d_m^{\gamma+1}c_{m,l}^{\gamma+1}}\,\|w_2^{\gamma+1}\Psi_{m,l}^{\gamma+1}\|_{\mu_3^\gamma} = \pi\sqrt{d_m^{\gamma+1}c_{m,l}^{\gamma+1}}\,\|\Psi_{m,l}^{\gamma+1}\|_{\mu_2^{\gamma+1}}\\ &= \pi\sqrt{d_m^{\gamma+1}c_{m,l}^{\gamma+1}},\end{aligned}$$

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

$$\|D^p\bar{D}^q z_{m,l}\|_{\mu_3^s} = \frac{s!}{\pi}\,\|h_{m-s,l-p}^{s+1}\|_{\mu_3^s} = s!\sqrt{d_{m-s}^{\gamma+1}c_{m-s,l-p}^{\gamma+1}}.$$

Plugging in the formulas for $c_{m,l}^\gamma$ and d_m^γ given above provides (55). □

B.3. 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 $\lambda \equiv 1$, 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 $\lambda \equiv 1$ 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 B.7 and using $\Gamma(m+2\gamma)/\Gamma(m+1) \sim m^{2\gamma-1}$,

$$\begin{aligned}\sum_{l=0}^m c_{m,l}^{-1} &= \frac{\Gamma(\gamma)^2}{\Gamma(2\gamma)}\frac{\Gamma(m+2\gamma)}{\Gamma(m+1)}\sum_{l=0}^m \frac{\Gamma(l+1)}{\Gamma(l+\gamma)}\frac{\Gamma(m-l+1)}{\Gamma(m-l+\gamma)}\\ &\sim C_\gamma m^2,\end{aligned}$$

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 A1 for $\rho = 2$ and $C = C_\gamma$, Theorem 3.6 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 B.7. *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

$$1-\varepsilon \le f(x+1)/x^{1-\gamma} \le 1+\varepsilon, \qquad x \ge x_\varepsilon. \tag{57}$$

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,

$$\begin{aligned}\sum_{l=m_0}^{m/2} f(l+1) f(m-l+1) &\gtrsim \sum_{l=m_0}^{m/2} l^{1-\gamma}(m-l)^{1-\gamma} \ge (m/2)^{1-\gamma} \sum_{l=m_0}^{m/2} l^{1-\gamma} \\ &\gtrsim m^{3-2\gamma}.\end{aligned}$$

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. $\square$