Ann. Inst. Statist. Math.
Vol. 53, No. 4, 781–798 (2001)

ABSTRACT INVERSE ESTIMATION WITH APPLICATION TO DECONVOLUTION ON LOCALLY COMPACT ABELIAN GROUPS *

ARNOUD C. M. VAN ROOIJ[1] AND FRITS H. RUYMGAART[2]

[1]*Department of Mathematics, Katholieke Universiteit Nijmegen, 6525 ED Nijmegen, The Netherlands*
[2]*Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, U.S.A.*

(Received December 22, 1997; revised June 19, 2000)

Abstract. Recovery of the unknown parameter in an abstract inverse estimation model can be based on regularizing the inverse of the operator defining the model. Such regularized-inverse type estimators are constructed with the help of a version of the spectral theorem due to Halmos, after suitable preconditioning. A lower bound to the minimax risk is obtained exploiting the van Trees inequality. The proposed estimators are shown to be asymptotically optimal in the sense that their risk converges to zero, as the sample size tends to infinity, at the same rate as this lower bound. The general theory is applied to deconvolution on locally compact Abelian groups, including both indirect density and indirect regression function estimation.

Key words and phrases: Abstract inverse estimation, indirect curve estimation, ill-posed problem, regularized-inverse type estimator, locally compact Abelian group, deconvolution.

1. Introduction and preliminaries

In this paper we consider the general abstract inverse estimation or indirect curve estimation model and, as an important special subclass, noisy convolution models on locally compact Abelian groups. Such groups contain $\mathbb{R}^d$, $(0,\infty)^d$, $\mathbb{Z}^d$, and $\mathbb{T}^d$ ($\mathbb{T}$ is the complex unit circle) for any finite dimension d as special cases. One may add groups like $\mathbb{Z}$ mod k to this list, but that would not exhaust the family of groups that can be constructed from the above. An interesting example is $\mathbb{Q}_p$ (p is an arbitrary prime number), the group of p-adic numbers. This nontrivial group is not discrete, not compact, and has nothing to do with $\mathbb{R}^d$, $\mathbb{Z}^d$, $\mathbb{T}^d$, or $\mathbb{Z}$ mod k. Although our methods remain valid for $\mathbb{Z}$ mod k and more exotic groups like $\mathbb{Q}_p$, they are not of sufficient practical importance to be included in our discusssion.

In the general case a random sample $X_1, \dots, X_n$ of independent copies of a random element X is given which is directly related to an element p of a separable Hilbert space $\mathbb{L}$. For convolutions this will be specified below. The element p is the image of an unknown element θ in a given subset Θ of a separable Hilbert space $\mathbb{H}$ under a known linear, bounded, and injective operator $K : \mathbb{H} \to \mathbb{L}$, i.e.

$$p = K\theta, \qquad \theta \in \Theta \subset \mathbb{H}. \tag{1.1}$$

*Research supported by NSF grants DMS-92-04950 and DMS-95-04485.

The problem is to recover θ (also called the input, input signal, or input function) from the data related to p (also called the output, output signal, or output function), and to assess the quality of such an estimator.

A very important category of inverse problems is given by the deconvolutions mentioned above. In that case we have $\mathbb{H} = \mathbb{L} = L^2(\mu_{\mathbb{G}})$, the Hilbert space of all functions on a locally compact Abelian group $\mathbb{G}$ that are square integrable with respect to the Haar measure $\mu_{\mathbb{G}}$, and the operator $K = K_w$ is the convolution with a kernel w that is assumed to satisfy

$$w \in L^1(\mu_{\mathbb{G}}) \cap L^2(\mu_{\mathbb{G}}). \tag{1.2}$$

Writing the group $\mathbb{G}$ additively with group operation $\oplus$ and reciprocal $\ominus$ we arrive at the model

$$\begin{aligned} p(x) &= \int_{\mathbb{G}} w(x \ominus y)\theta(y)d\mu_{\mathbb{G}}(y) \\ &=: (w \circledast \theta)(x) =: (K_w\theta)(x), \qquad x \in \mathbb{G}, \quad \theta \in \Theta \subset L^2(\mu_{\mathbb{G}}). \end{aligned} \tag{1.3}$$

The assumption that w is also square integrable with respect to $\mu_{\mathbb{G}}$ is to ensure the existence of the mean integrated square error (MISE) of our estimators.

In the theory of integral equations preconditioning is an important concept. It will be applied here in the sense that (1.1) will be replaced with the equivalent equation

$$q := K^*p = K^*K\theta =: R\theta, \qquad \theta \in \Theta, \tag{1.4}$$

where K^* is the adjoint of K. In the case of convolutions K_w^* is convolution with $w^*(x) := w(\ominus x)$, $x \in \mathbb{G}$, and hence

$$q = K_r\theta = r \circledast \theta, \qquad \text{where} \quad r := w^* \circledast w. \tag{1.5}$$

Preconditioning has the advantage that the operator R is somewhat simpler than K, since it is strictly positive Hermitian. A further advantage is that q is usually easier to estimate from the data than p. This is because q is a smoothed version of p which allows an unbiased and $\sqrt{n}$-consistent estimator. Here follow two basic examples.

Example 1.1. Suppose that Z and E are independent random variables in $\mathbb{G}$, where E has known density w with respect to $\mu_{\mathbb{G}}$ and the density θ of Z with respect to this measure is unknown. We observe n independent copies of

$$X := Z \oplus E, \tag{1.6}$$

where X has density $p = w \circledast \theta$. Clearly an unbiased and $\sqrt{n}$-consistent estimator of $q = w^* \circledast p$ is given by

$$\hat{q}(x) := \frac{1}{n}\sum_{k=1}^{n} w^*(x \ominus X_k) = \frac{1}{n}\sum_{k=1}^{n} w(X_k \ominus x) =: \frac{1}{n}\sum_{k=1}^{n} \hat{q}_k(x), \qquad x \in \mathbb{G}. \tag{1.7}$$

For the unbiasedness just observe that $\boldsymbol{E}w^*(\bullet \ominus X) = w^* \circledast p = r \circledast \theta = q$ in the notation of (1.5). Since $\boldsymbol{E}\|w^*(\bullet \ominus X)\|^2 = \boldsymbol{E}\int_{\mathbb{G}} |w^*(x \ominus X)|^2 d\mu_{\mathbb{G}}(x) = \int_{\mathbb{G}} |w^*(y)|^2 d\mu_{\mathbb{G}}(y) =$

$\|w\|^2 < \infty$ thanks to assumption (1.2), it follows that $\boldsymbol{E}\|\hat{q}\|^2 < \infty$. Jointly with the unbiasedness this entails

$$
\begin{aligned}
(1.8) \qquad P\{\sqrt{n}\|\hat{q}-q\| \geq C\} &\leq \frac{n}{C^2}\boldsymbol{E}\|\hat{q}-q\|^2 \\
&= \frac{1}{C^2}\int_{\mathbb{G}} \operatorname{Var}(w^*(x\ominus X))d\mu_{\mathbb{G}}(x) \leq \left(\frac{\|w\|}{C}\right)^2 \to 0, \qquad \text{as} \quad C\to\infty,
\end{aligned}
$$

which settles the $\sqrt{n}$-consistency.

Example 1.2. Let us now assume that Z is a random variable in $\mathbb{G}$ with density f_1 with respect to $\mu_{\mathbb{G}}$, and that E is a real valued random variable, independent of Z, with density f_2 with respect to Lebesgue measure that has mean 0 and finite variance σ^2. We observe n independent copies of

$$
(1.9) \qquad X := (Y, Z), \qquad Y := (w \circledast \theta)(Z) + E = p(Z) + E.
$$

The joint density of Y and Z is given by

$$
(1.10) \qquad f(y,z) = f_2(y - (w\circledast\theta)(z))f_1(z), \qquad y\in\mathbb{R}, \quad z\in\mathbb{G}.
$$

An unbiased estimator of $q = w^* \circledast p$ is now given by

$$
(1.11) \qquad \hat{q}(x) := \frac{1}{n}\sum_{k=1}^{n}\frac{Y_k}{f_1(Z_k)}w^*(x\ominus Z_k) =: \frac{1}{n}\sum_{k=1}^{n}\hat{q}_k(x), \qquad x\in\mathbb{G}.
$$

To see this note that $\boldsymbol{E}(Y/f_1(Z))w^*(\bullet\ominus Z) = \boldsymbol{E}\boldsymbol{E}((Y/f_1(Z))w^*(\bullet\ominus Z)\mid Z) = \boldsymbol{E}(w^*(\bullet\ominus Z)/f_1(Z))\boldsymbol{E}(Y\mid Z) = \boldsymbol{E}w^*(\bullet\ominus Z)(w\circledast\theta)(Z)/f_1(Z) = \int_{\mathbb{G}} w^*(\bullet\ominus z)(w\circledast\theta)(z)d\mu_{\mathbb{G}}(z) = r\circledast\theta = q$.

Again we have $\boldsymbol{E}\|\hat{q}\|^2 < \infty$, but now we need the condition.

$$
(1.12) \qquad m := \operatorname{ess}\inf_{z\in\mathbb{G}} f_1(z) > 0.
$$

Of course this condition can only be satisfied if $\mathbb{G}$ has finite Haar measure (like, e.g., the torus $\mathbb{T}^d$). If this condition is not fulfilled the procedure has to be slightly modified where the unbiasedness of $\hat{q}$ has to be sacrificed. Since all functions are real and E has zero mean and variance σ^2, we have

$$
\begin{aligned}
(1.13) \quad \boldsymbol{E}\left\|\frac{Y}{f_1(Z)}w^*(\bullet\ominus Z)\right\|^2 &= \int_{\mathbb{G}}\boldsymbol{E}\left|\frac{Y}{f_1(Z)}w^*(\bullet\ominus Z)\right|^2 d\mu_{\mathbb{G}} \\
&= \int_{\mathbb{G}}\boldsymbol{E}\boldsymbol{E}\left(\frac{(w\circledast\theta)^2(Z) + 2E(w\circledast\theta)(Z) + E^2}{f_1^2(Z)}\{w^*(\bullet\ominus Z)\}^2 \,\middle|\, Z\right)d\mu_{\mathbb{G}} \\
&= \int_{\mathbb{G}}\boldsymbol{E}\frac{\{(w\circledast\theta)^2(Z) + E^2\}\{w^*(\bullet\ominus Z)\}^2}{f_1^2(Z)}d\mu_{\mathbb{G}} \\
&= \boldsymbol{E}\frac{(w\circledast\theta)^2(Z) + E^2}{f_1^2(Z)}\cdot\int\{w^*(\bullet\ominus Z)\}^2 d\mu_{\mathbb{G}} \\
&\leq \frac{1}{m}\left\{\int_{\mathbb{G}}(w\circledast\theta)^2 d\mu_{\mathbb{G}} + \boldsymbol{E}\frac{E^2}{f_1(Z)}\right\}\|w\|^2 = \frac{1}{m}(\|p\|^2+\sigma^2)\|w\|^2.
\end{aligned}
$$

This yields at once the $\sqrt{n}$-consistency as in (1.8). This model is a random design indirect regression model. Extensions to the case where f_1 is unknown and has to be estimated from the data, or the design is deterministic are possible but will not be considered here.

Further examples can be found in Koo (1993). Practical examples include for instance

- curve estimation for directional data (Mardia and Jupp (2000));
- recovering the initial heat distribution from the present state (Margenau and Murphy (1956), Mair and Ruymgaart (1995));
- recovering the input function of a dynamical system (Gilliam *et al.* (1988), Chauveau *et al.* (1994), Dey *et al.* (1998));
- errors in variables models (Carroll and Hall (1988), Zhang (1990), Fan (1991));
- image restoration models (Hall (1990), Donoho (1994)).

We will now briefly review the general construction of estimators of regularized-inverse (RI) type. According to Halmos' (1963) version of the spectral theorem there exists a σ-finite measure space $(\mathbb{S}, \mathcal{S}, \nu)$, a unitary operator $U : \mathbb{H} \to L^2(\nu)$, and a function $\rho \in L^\infty(\nu)$ which is strictly positive ν-a.e., such that

$$R = U^{-1} M_\rho U, \tag{1.14}$$

employing the notation M_ρ for the operator acting on $L^2(\nu)$ as multiplication with ρ. The exact inverse of R is unitarily equivalent with division by ρ. More specifically, this means that $\theta = R^{-1}q = U^{-1}(1/\rho)Uq$, see (1.4). In general R^{-1} is unbounded, so that the estimator $\hat{q}$ of q may not be in the range of the inverse and, if it is, $R^{-1}\hat{q}$ may not be close to $R^{-1}q$. Therefore a regularized version of the inverse will be used. Let us consider the functions

$$\delta_\alpha(t) := \frac{1}{t}\mathbf{1}_{[\alpha,\infty)}(t), \qquad t > 0, \tag{1.15}$$

for $\alpha > 0$. A family of regularized inverses is now given by

$$R_\alpha^{-1} := \delta_\alpha(R) := U^{-1} M_{\delta_\alpha(\rho)} U, \qquad \alpha > 0. \tag{1.16}$$

Note that $R_\alpha^{-1} : \mathbb{H} \to \mathbb{H}$ is bounded for each $\alpha > 0$, and that $\|R_\alpha^{-1}Rf - f\| \to 0$, as $\alpha \downarrow 0$, for each $f \in \mathbb{H}$.

The *RI*-estimators, proposed here, are defined as

$$\hat{\theta}_\alpha := R_\alpha^{-1}\hat{q}, \qquad \text{for suitable} \quad \alpha > 0. \tag{1.17}$$

What "suitable" means will be discussed in Sections 2 and 3. For convolution models the construction can be made explicit as we will see below in Section 3. The regularization scheme (1.15) employed throughout this paper is called spectral cut-off regularization (Kress (1989)). Another well-known possibility is Tikhonov type regularization based on the family of functions $t \mapsto 1/(\alpha + t)$, $t > 0$, $\alpha > 0$ (Tikhonov and Arsenin (1977), Vapnik(1982)). In finite dimensional parametric regression the latter scheme yields ridge regression as introduced by Hoerl and Kennard(1970*a*, 1970*b*).

An important aspect of curve estimation relates to convergence rates of $\boldsymbol{E}\|\hat{\theta}_\alpha - \theta\|^2$, the MISE. In a seminal paper Pinsker (1980) derived a lower bound which is asymptotically attainable at the level of constants. An extension to adaptive estimation can be found in Efromovich and Pinsker (1984). Pinsker (1980) emphasizes a model of direct estimation of an infinite dimensional parameter with Gaussian errors. Extensions to certain indirect estimation models can be found in the literature. Such results, however, often concern models with compact operators like in Johnstone and Silverman (1990), where the compact Radon transform is emphasized. Also deconvolution on the circle $\mathbb{T}$ can be found. This convolution is compact in contrast to most other convolutions. Extension to the general bounded operators considered here doesn't seem immediate. In order to apply Pinsker's (1980) result, moreover, many authors assume that a signal is—directly or indirectly—given in the presence of Gaussian white noise. This assumption is justified for curve estimation models that are asymptotically equivalent with such a model like for univariate regression (Brown and Low (1996)) or univariate density estimation (Nussbaum (1996)). An alternative approach is provided by extensions of Pinsker's (1980) inequality to non-Gaussian errors; see, for instance, Golubev and Nussbaum (1990). For an interesting review of the Pinsker bound and its ramifications see Nussbaum (1996).

Here we first elaborate on a lower bound in van Rooij and Ruymgaart (1996) valid for the general model (1.4), that was obtained in a quite different way by coordinatewise application of the van Trees inequality (Gill and Levit (1995)). It turns out to be possible to represent this lower bound as an integral with respect to the measure ν on the spectral domain $\mathbb{S}$. Since an upper bound can also be represented as an integral with respect to that measure, it is possible to arrive at a comparison between the two and to establish optimality of the rate of convergence of the MISE of the estimators in (1.17), provided that the regularization parameter $\alpha = \alpha_n \to 0$, as $n \to \infty$, at a suitable rate. To the best of our knowledge at the present level of generality this optimality result is new. Both the upper and the lower bound are considered in Section 2. The results are applied to deconvolution in Section 3, where we show in particular that in this case a crucial condition for attainment of the optimal rate is fulfilled. Some specific examples are considered in Section 4. The first of these is an instance where deconvolution is not ill-posed.

It is a common problem in nonparametric curve estimation that such rates depend on the presupposed smoothness of the input function. Bowman (1984) and Rudemo (1982) proposed a cross-validation method for data-driven selection of the smoothness parameter. This method was generalized to the general indirect curve estimation model (1.4) in Dey *et al.* (1996) and applied to deconvolution on the real line with good results. In Section 3 we also specify this cross-validation method for abstract convolutions. Other methods to deal with adaptation and inhomogeneous smoothness of the function to be estimated are based on wavelet-vaguelette decompositions (Donoho (1995)), and on the linear functional strategy used for ill-posed inverse problems (Goldenshluger (1997)).

A great variety of practical examples is related to convolution. Such examples as tomography (Johnstone and Silverman (1990)) or Wicksell's problem (Groeneboom and Jongbloed (1995), Nychka and Cox (1989)) are not directly related to convolution, although they fit in the general framework (Caroll *et al.* (1991)). There are interesting inverse estimation problems of deconvolution type on homogeneous spaces that are not Abelian groups. The problem of deconvolution on the sphere has been proposed in van Rooij and Ruymgaart (1991) and solved by Healy and Kim (1993, 1996), see also Healy *et al.* (1998). The latter authors also provide an interesting application to geometric quality assurance. Kim (1998) considers deconvolution on the special orthogonal group. Another example is the Poincaré upper half plane. Although for this space Fourier theory seems available (Terras (1985)) specific technicalities will arise. Therefore this example will not be considered here. For many other interesting statistical models in non-Euclidean spaces see Diaconis (1988).

Also the question of proper dealing with prior knowledge about the presence or lack of smoothness of the input signal will not be considered here. The first can be dealt with by adopting methods from Mair and Ruymgaart (1995) and by embedding the model in a Sobolev scale. If the input signal is known to be irregular, modifications will be required to control the Gibbs phenomenon, and the MISE may no longer be suitable to assess the quality of estimators. One might rather employ the Hausdorff-distance between the closed graphs of estimator and estimand as suggested by Marron and Tsybakov (1995). This idea was carried out in Chandrawansa *et al.* (1999) in conjunction with Cesàro-averaging. See also Neumann (1995) for indirect estimation of change points.

2. Asymptotic optimality for general abstract estimators

It has already been observed that it will in general be possible to estimate q in (1.4) unbiasedly and $\sqrt{n}$-consistently. Such an estimator will be of the form $\hat{q} := n^{-1}\sum_{k=1}^{n}\hat{q}_k$, where the $\hat{q}_k$ are i.i.d. random elements in $\mathbb{H}$ with $\hat{q}_k$ depending on X_k only, and we assume they satisfy

$$\boldsymbol{E}\|\hat{q}_k\|^2 < \infty, \qquad \boldsymbol{E}\hat{q}_k = q. \tag{2.1}$$

Examples of such estimators are given in (1.7) and (1.11).

The estimators $\hat{q}$ lead to the estimators $\hat{\theta}_\alpha$ in (1.17) of actual interest. In this section we will establish the asymptotic optimality of those estimators under the assumption that there exists a constant $0 < C < \infty$ such that

$$\operatorname{Var}(U\hat{q})(s) \le \frac{C}{n}\rho(s), \qquad s \in \mathbb{S}, \tag{2.2}$$

where ρ is the function that represents the operator R in the spectral domain as reviewed in Section 1. The assumption implies that preconditioning (see (1.4)) does not introduce extra, undue, ill-posedness. It is typically fulfilled if preconditioning is carried out with the adjoint operator, and in Section 4 conditions will be specified under which it is satisfied for the estimators in (1.7) and (1.11).

It is clear that under this assumption

$$\boldsymbol{E}\|\hat{\theta}_\alpha - \theta\|^2 \le \frac{C}{n}\int_{\{\rho\ge\alpha\}}\frac{1}{\rho}d\nu + \int_{\{\rho<\alpha\}}|U\theta|^2 d\nu, \tag{2.3}$$

see van Rooij and Ruymgaart ((1996), Theorem 2.1). Since $\theta \in \mathbb{H}$ it follows that $U\theta \in L^2(\nu)$ and therefore, without assuming anything more about θ, we still can prove from (2.3) the existence of a sequence $\alpha(n) \downarrow 0$, as $n \to \infty$, such that the $\hat{\theta}_{\alpha(n)}$ are consistent estimators of θ. For a speed of convergence of the MISE we need to assume more about θ. In concrete cases usually smoothness assumptions are made. Such assumptions may follow from the abstract setting by bounding $|U\theta|$. For $\ell \in L^2(\nu)$ let us consider the set

$$\Theta_0 := \{\theta \in \mathbb{H} : |U\theta| \le \ell\}; \qquad \text{we assume} \quad \Theta_0 \subset \Theta. \tag{2.4}$$

It should be noted that such submodels are designed to describe smoothness classes, but may include irregular functions if ℓ has relatively heavy tails. The following result is immediate from (2.3).

THEOREM 2.1. *For the submodel Θ_0 satisfying* (2.4) *we have*

$$\sup_{\theta\in\Theta_0} \boldsymbol{E}\|\hat{\theta}_\alpha - \theta\|^2 \le \frac{C}{n}\int_{\{\rho\ge\alpha\}} \frac{1}{\rho} d\nu + \int_{\{\rho<\alpha\}} \ell^2 d\nu, \tag{2.5}$$

provided that (2.1) *and* (2.2) *are fulfilled.*

The construction and attainment of the lower bound require a number of further specifications and assumptions. Let us take $\ell \in L^\infty(\nu) \cap L^2(\nu)$ real valued with $\ell_0 \ge$ ess sup ℓ and $\ell_0 > 0$, strictly positive numbers $\ell_0 > \ell_1 > \ell_2 > \cdots \downarrow 0$, and sets $S_k := \{s \in \mathbb{S} : \ell_k \le \ell(s) < \ell_{k-1}\}$. We assume that there exist $0 < a < \infty$ and $0 < \underline{\nu} \le \bar{\nu} < \infty$ such that

$$\ell_k \ge a\ell_{k-1}, \quad 0 < \underline{\nu} \le \nu_k := \nu(S_k) \le \bar{\nu} < \infty, \quad k \in \mathbb{N}. \tag{2.6}$$

Let us introduce the orthonormal system $\varphi_k := (\nu_k)^{-1/2}\mathbf{1}_{S_k}$ in $L^2(\nu)$ and note that the $e_k := U^{-1}\varphi_k$ form an orthonormal system in $\mathbb{H}$. Also observe that

$$\begin{aligned}\mathcal{T}_0 &:= \{\theta \in \mathbb{H} : U\theta = \sum_{k=1}^{\infty} t_k\varphi_k : |t_k| \le \ell_k\sqrt{\nu_k}\} \\ &= \left\{\theta \in \mathbb{H} : \theta = \sum_{k=1}^{\infty} t_k e_k : |t_k| \le \ell_k\sqrt{\nu_k}\right\} \subset \Theta_0,\end{aligned} \tag{2.7}$$

where the inclusion is obvious from $\ell \ge \sum_{k=1}^{\infty} \ell_k \mathbf{1}_{S_k} = \sum_{k=1}^{\infty} \ell_k\sqrt{\nu_k}\varphi_k$. It is convenient to identify $\mathcal{T}_0$ with $I_0 := [-\ell_1\sqrt{\nu_1}, \ell_1\sqrt{\nu_1}] \times [-\ell_2\sqrt{\nu_2}, \ell_2\sqrt{\nu_2}] \times \cdots$ and $\theta \in \mathcal{T}_0$ with $t = (t_1, t_2, \ldots)$.

Next let us be more explicit about the random sample $X_1, \ldots, X_n$ and assume that the X_k take values in a σ - finite measure space $(\mathbb{X}, \mathcal{X}, \mu)$ with common density f_t, $t \in I_0$, with respect to μ. For each $t \in I_0$ the $\sqrt{f_t}$ are supposed to be strongly differentiable in $L^2(\mu)$ with respect to $t_k \in [-\ell_k\sqrt{\nu_k}, \ell_k\sqrt{\nu_k}]$. The derivative will be denoted by $\partial\sqrt{f_t}/\partial t_k$. Regarding the Fisher information we assume that

$$\sup_{t\in I_0} \|\partial\sqrt{f_t}/\partial t_k\|^2 \le \rho_k, \quad k \in \mathbb{N}, \tag{2.8}$$

for finite numbers ρ_k such that

$$S_k \subset \{s \in \mathbb{S} : \rho(s) \ge b\rho_k\}, \quad k \in \mathbb{N}, \tag{2.9}$$

for some $0 < b < \infty$.

THEOREM 2.2. *Let assumptions* (2.4), (2.6), (2.8), *and* (2.9) *be fulfilled. Then we have*

$$\inf_{T\in\mathcal{E}} \sup_{\theta\in\Theta_0} \boldsymbol{E}\|T - \theta\|^2 \ge \int \frac{\ell^2(s)}{A + Bn\rho(s)\ell^2(s)} d\nu(s), \tag{2.10}$$

where $\mathcal{E}$ is the class of all $\mathbb{H}$ - valued estimators T with finite MISE, and where A, B $\in (0,\infty)$ are constants.

PROOF. In van Rooij and Ruymgaart (1996) it has been shown that (2.4) and (2.8) suffice to obtain

$$\inf_{T\in\mathcal{E}}\sup_{\theta\in\Theta_0} E\|T-\theta\|^2 \geq \sum_{k=1}^{\infty}\frac{\ell_k^2}{c+n\rho_k\ell_k^2}, \tag{2.11}$$

for some $0 < c < \infty$, by exploiting the van Trees inequality (see Gill and Levit (1995)). It remains to relate the summation in (2.11) to the integral in (2.10). Now we use (2.6) and (2.9) to obtain

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{\ell_k^2}{c+n\rho_k\ell_k^2} &\geq \frac{1}{\bar\nu}\sum_{k=1}^{\infty}\int_{S_k}\frac{1}{(c/\ell_k^2)+n\rho_k}d\nu \\ &\geq \frac{1}{\bar\nu}\sum_{k=1}^{\infty}\int_{S_k}\frac{1}{(c/a^2\ell^2)+(n/b)\rho}d\nu \\ &= \sum_{k=1}^{\infty}\int_{S_k}\frac{\ell^2}{(\bar\nu c/a^2)+(\bar\nu n/b)\rho\ell^2}d\nu,\end{aligned} \tag{2.12}$$

which equals the integral on the right in (2.10) with $A = \bar\nu c/a^2$ and $B = \bar\nu/b$.

To verify that the RI-estimators actually attain the minimax rate over Θ_0 we need a further assumption which relates ℓ to ρ. The function ρ is imposed on us, but we are more or less free in the choice of ℓ. In concrete cases ℓ should decay to zero in the tails and our objective for ℓ can therefore be often realized by setting

$$\ell = \Psi(\rho), \tag{2.13}$$

for suitable $\Psi : [0,\infty) \to [0,\infty)$, which is assumed to be continuous and strictly increasing with $\Psi(0) = 0$.

THEOREM 2.3. *Under assumptions* (2.1), (2.2), (2.4), (2.6), (2.8), (2.9), *and* (2.13) *it is possible to choose* $\alpha = \alpha(n) > 0$ *in such a way that the MISE of the RI-estimators* $\hat\theta_{\alpha(n)}$ *in* (1.17) *converges to* 0 *at the minimax rate, i.e. at the rate of the integral on the right in* (2.10), *as* $n \to \infty$.

PROOF. For n sufficiently large it is possible to choose $\alpha = \alpha(n) > 0$ such that $\alpha\Psi^2(\alpha) = A/Bn$. For this α we then have

$$\left[\rho \begin{smallmatrix}\geq\\=\\<\end{smallmatrix} \alpha\right] \Rightarrow \left[\rho\Psi^2(\rho) \begin{smallmatrix}\geq\\=\\<\end{smallmatrix} \frac{A}{Bn}\right] \Rightarrow \left[Bn\rho \begin{smallmatrix}\geq\\=\\<\end{smallmatrix} \frac{A}{\ell^2}\right], \tag{2.14}$$

using (2.13). It follows that for each n (sufficiently large)

$$\begin{aligned}\int_{\mathbb{S}}\frac{\ell^2}{A+Bn\rho\ell^2}d\nu &= \int_{\{\rho>0\}}\frac{1}{A/\ell^2+Bn\rho}d\nu \\ &= \left(\int_{\{\rho\geq\alpha\}}+\int_{\{0<\rho<\alpha\}}\right)\frac{1}{A/\ell^2+Bn\rho}d\nu\end{aligned} \tag{2.15}$$

$$\geq \int_{\{\rho\geq\alpha\}} \frac{1}{2Bn\rho} d\nu + \int_{\{0<\rho<\alpha\}} \frac{1}{2A/\ell^2} d\nu$$
$$= \frac{1}{2Bn} \int_{\{\rho\geq\alpha\}} \frac{1}{\rho} d\nu + \frac{1}{2A} \int_{\{\rho<\alpha\}} \ell^2 d\nu$$
$$\geq \left(\frac{1}{2A} \wedge \frac{1}{2BC}\right) \left\{\frac{C}{n} \int_{\{\rho\geq\alpha\}} \frac{1}{\rho} d\nu + \int_{\{\rho<\alpha\}} \ell^2 d\nu\right\}.$$

It should be noted that the number $(1/2A) \wedge (1/2BC)$ does not depend on n. (Combining Theorem 2.1 and Theorem 2.2 we actually see that this number cannot be larger than 1.) Since, of course, the right hand side in (2.5) is at least as great as the lower bound on the right in (2.10) (for any $\alpha > 0$) the theorem follows.

3. Application to convolution models

The results will now be applied to abstract convolution and more in particular to the indirect density and regression estimation models in Examples 1.1 and 1.2. We will briefly review some basic facts of abstract harmonic analysis and refer to Hewitt and Ross (1963) for further information. A character γ of $\mathbb{G}$ is a mapping $\gamma : \mathbb{G} \to \mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ which is a continuous homomorphism, meaning that γ is continuous and $\gamma(x \oplus y) = \gamma(x) \cdot \gamma(y)$ for $x, y \in \mathbb{G}$. These characters form an l.c. Abelian group Γ under pointwise multiplication. Haar measure on Γ is denoted by μ_Γ, and $L^p(\mu_\Gamma)$ is defined in the usual way. (It should be noted that only in exceptional cases $\mathbb{G}$ and Γ are isomorphic: in fact $\mathbb{G} = \mathbb{R}^d$ is such a case.)

The unitary operator that reduces convolutions to multiplications is the Fourier transform. The Fourier transform of $w \in L^1(\mu_\mathbb{G})$ is the mapping $\tilde{w} : \Gamma \to \mathbb{C}$ defined by

$$\tilde{w}(\gamma) := \int_\mathbb{G} w(x)\gamma(x) d\mu_\mathbb{G}(x), \qquad \gamma \in \Gamma. \tag{3.1}$$

It can be shown that a linear isometry $\mathcal{F}_\mathbb{G} : L^2(\mu_\mathbb{G}) \to L^2(\mu_\Gamma)$ exists, called the Fourier-Plancherel transform, such that

$$\mathcal{F}_\mathbb{G} w = c_\mathbb{G} \tilde{w}, \qquad w \in L^1(\mu_\mathbb{G}) \cap L^2(\mu_\mathbb{G}), \tag{3.2}$$

for a certain number $c_\mathbb{G}$. One can, moreover, show that $\mathcal{F}_\mathbb{G}(w \circledast \theta) = \tilde{w}.\ \mathcal{F}_\mathbb{G}\theta$, for $w \in L^1(\mu_\mathbb{G}) \cap L^2(\mu_\mathbb{G})$, so that by continuity

$$K_w = \mathcal{F}_\mathbb{G}^{-1} M_{\tilde{w}} \mathcal{F}_\mathbb{G}, \qquad w \in L^1(\mu_\mathbb{G}), \tag{3.3}$$

where $M_{\tilde{w}} : L^2(\mu_\Gamma) \to L^2(\mu_\Gamma)$ is the multiplication operator defined by $M_{\tilde{w}}\varphi := \tilde{w} \cdot \varphi, \varphi \in L^2(\mu_\Gamma)$, and K_w is convolution with w (see (1.3)).

The l.c. Abelian group Γ has its own character group that can be identified with $\mathbb{G}$. Hence there also exists a Fourier-Plancherel transform $\mathcal{F}_\Gamma : L^2(\mu_\Gamma) \to L^2(\mu_\mathbb{G})$. This transform is again an isometry and it can be shown that

$$\mathcal{F}_\Gamma \mathcal{F}_\mathbb{G} = S, \qquad \text{or} \qquad \mathcal{F}_\mathbb{G}^{-1} = S\mathcal{F}_\Gamma, \tag{3.4}$$

where $S : L^2(\mu_\mathbb{G}) \to L^2(\mu_\mathbb{G})$ is the reflection operator $(S\theta)(x) := \theta(\ominus x)$, $x \in \mathbb{G}$, with S^2 obviously equal to the identity.

Example 3.1. Let $\mathbb{G} = \mathbb{R}^d$ with $a \oplus b = a + b (a, b \in \mathbb{R}^d)$ and Haar measure equal to Lebesgue measure. Then the character corresponding to $x \in \mathbb{R}^d$ is the function $t \mapsto \exp(i\langle x, t\rangle), t \in \mathbb{R}^d$. We have indeed, that $x + y$ corresponds to $\exp(i\langle x+y, \bullet\rangle) = \exp(i\langle x, \bullet\rangle)\exp(i\langle y, \bullet\rangle)$. It follows that Γ is also $\mathbb{R}^d$ with Lebesgue measure as an invariant measure, and

$$(3.5) \qquad (\mathcal{F}_{\mathbb{R}^d} w)(t) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\langle x,t\rangle} w(x)dx, \qquad t \in \mathbb{R}^d,$$

for $w \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Of course $\mathcal{F}_\Gamma = \mathcal{F}_\mathbb{G}$ and it follows from applying (3.4) that

$$(3.6) \qquad (\mathcal{F}_{\mathbb{R}^d}^{-1} \varphi)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\langle x,t\rangle} \varphi(t)dt, \qquad x \in \mathbb{R}^d.$$

Example 3.2. Let $\mathbb{G} = (0,\infty)^d$ with $a \oplus b = a \cdot b (a, b \in (0,\infty)^d$, multiplication coordinatewise), and Haar measure being determined by its density $x \mapsto x^{-1} = \Pi_{j=1}^d x_j^{-1}$, $x \in (0,\infty)^d$, with respect to Lebesgue measure. In this case it turns out that $\Gamma = \mathbb{R}^d$ with Lebesgue measure as invariant measure. For $w \in L^1(\mu_{(0,\infty)^d}) \cap L^2(\mu_{(0,\infty)^d})$ we find

$$(3.7) \quad (\mathcal{F}_{(0,\infty)^d} w)(t) = (2\pi)^{-d/2} \int_{(0,\infty)^d} e^{i\sum_{j=1}^d t_j \log x_j} w(x) \Pi_{j=1}^d x_j^{-1} dx, \qquad t \in \mathbb{R}^d.$$

Formula (3.4) yields in this case ($\varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$)

$$(3.8) \qquad (\mathcal{F}_{(0,\infty)^d}^{-1} \varphi)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\sum_{j=1}^d t_j \log x_j} \varphi(t)dt, \qquad x \in (0,\infty)^d.$$

Example 3.3. Next let us take $\mathbb{G} = \mathbb{Z}^d$ with $m \oplus n = m + n (m, n \in \mathbb{Z}^d)$ and the counting measure as Haar measure. In this example $\Gamma = \mathbb{T}^d := \{u \in \mathbb{C}^d : |u_1| = \cdots = |u_d| = 1\}$, which can be identified with $[0, 2\pi)^d$ under addition mod 2π and with Lebesgue measure as invariant measure. We now have ($w \in \ell_d^1 \cap \ell_d^2$)

$$(3.9) \qquad (\mathcal{F}_{\mathbb{Z}^d} w)(t) = \sum_{n \in \mathbb{Z}^d} w(n) e^{i\langle n,t\rangle}, \qquad t \in [0, 2\pi)^d.$$

For $\varphi \in L^1([0,2\pi)^d) \cap L^2([0,2\pi)^d)$ application of (3.4) yields

$$(3.10) \qquad (\mathcal{F}_{\mathbb{Z}^d}^{-1} \varphi)(n) = \int_{[0,2\pi)^d} e^{-i\langle t,n\rangle} \varphi(t)dt, \qquad n \in \mathbb{Z}^d.$$

Example 3.4. This example follows from the previous one since $\mathbb{Z}^d$ and $\mathbb{T}^d$ are character group of one another. We have in particular

$$(3.11) \qquad (\mathcal{F}_{\mathbb{T}^d} w)(n) = \int_{[0,2\pi)^d} w(t) e^{i\langle n,t\rangle} dt, \qquad n \in \mathbb{Z}^d,$$

for $w \in L^1([0,2\pi)^d) \cap L^2([0,2\pi)^d)$, and for $\varphi \in \ell_d^1 \cap \ell_d^2$

$$(3.12) \qquad (\mathcal{F}_{\mathbb{T}^d}^{-1}\varphi)(t) = \sum_{n\in\mathbb{Z}^d} \varphi(n)e^{-i\langle t,n\rangle}, \qquad t \in [0,2\pi)^d.$$

Two simple conditions will be imposed on w, viz.

$$(3.13) \qquad \tilde{w} \neq 0 \text{ on } \Gamma, \qquad w \in L^1(\mu_{\mathbb{G}}) \cap L^2(\mu_{\mathbb{G}}).$$

The first entails the injectivity of K_w and also

$$(3.14) \qquad \tilde{r} = \tilde{w} \cdot \widetilde{w^*} = |\tilde{w}|^2 > 0 \qquad \text{on} \quad \Gamma.$$

Because of (3.14) the operator R is strictly positive Hermitian. We observed already that the second condition in (3.13) guarantees that the MISE of the estimators is well defined, as will be seen below.

The inverse R^{-1} is in general unbounded and therefore unsuitable for the construction of the estimators. In order to define a regularized version of this inverse we follow the pattern described in (1.15) and (1.16). A family of regularized inverses is now given by

$$(3.15) \qquad R_\alpha^{-1} := \mathcal{F}_{\mathbb{G}}^{-1} M_{\delta_\alpha(\tilde{r})} \mathcal{F}_{\mathbb{G}}, \qquad \alpha > 0,$$

where $M_{\delta_\alpha(\tilde{r})}$ is multiplication by $\delta_\alpha(\tilde{r})$ in $L^2(\mu_\Gamma)$. Let us write $\| \bullet \|_{\mathbb{G}}$ and $\| \bullet \|_\Gamma$ for the norm in $L^2(\mu_{\mathbb{G}})$ and $L^2(\mu_\Gamma)$ respectively. Note that $R_\alpha^{-1} : L^2(\mu_{\mathbb{G}}) \to L^2(\mu_{\mathbb{G}})$ is bounded for each $\alpha > 0$, and that

$$(3.16) \qquad \lim_{\alpha \downarrow 0} \|R_\alpha^{-1} Rw - w\|_{\mathbb{G}} = 0, \qquad \text{for each} \quad w \in L^2(\mu_{\mathbb{G}}),$$

according to the dominated convergence theorem.

The RI-estimators of θ are defined as

$$(3.17) \qquad \hat{\theta}_\alpha := R_\alpha^{-1}\hat{q}, \qquad \alpha > 0, \qquad \text{where} \quad \hat{q} := \frac{1}{n}\sum_{k=1}^{n} \hat{q}_k,$$

and where we assume that the $\hat{q}_k$ are i.i.d. with mean q, so that $\hat{q}$ is an unbiased estimator of q. Both in the case of indirect density and in the case of indirect regression function estimation, to be considered below, such an estimator $\hat{q}$ exists. Exploiting (3.15) the expression for $\hat{\theta}_\alpha$ can be given a more explicit form. In fact, it can be shown that the sets $\{\gamma : \tilde{r}(\gamma) \geq \alpha\}$, $\alpha > 0$, are compact and that $\delta_\alpha(\tilde{r}) \in L^1(\mu_\Gamma) \cap L^2(\mu_\Gamma)$; see Carroll *et al.* (1991). Using the obvious relation $\mathcal{F}_{\mathbb{G}}^{-1}(\delta_\alpha(\tilde{r}) \cdot \mathcal{F}_{\mathbb{G}}\hat{q}) = c_{\mathbb{G}}(\mathcal{F}_{\mathbb{G}}^{-1}\delta_\alpha(\tilde{r})) \circledast \hat{q}$, and (3.4) we arrive at

$$(3.18) \qquad \hat{\theta}_\alpha = \Delta_\alpha \circledast \hat{q}, \qquad \alpha > 0,$$

where $\Delta_\alpha(x) := c_\Gamma c_{\mathbb{G}} \int_\Gamma \delta_\alpha(\tilde{r}(\gamma))\overline{\gamma(x)}d\mu_\Gamma(\gamma)$.

To construct the estimator of θ we need the empirical character

$$(3.19) \qquad \hat{\gamma}(\gamma) := \frac{1}{n}\sum_{k=1}^{n} \gamma(X_k),$$

to deal with the model of Example 1.1, and

$$\hat{\chi}(\gamma) := \frac{1}{n}\sum_{k=1}^{n} \frac{Y_k}{f_1(Z_k)}\gamma(Z_k), \tag{3.20}$$

in order to deal with the model of Example 1.2. Recall that we need

$$w \in L^1(\mu_{\mathbb{G}}) \cap L^2(\mu_{\mathbb{G}}), \qquad |\tilde{w}| > 0, \tag{3.21}$$

in order to insure existence of the MISE and injectivity respectively.

THEOREM 3.1. *The RI-type indirect density estimator can be written as*

$$\hat{\theta}_\alpha = c_{\mathbb{G}}\mathcal{F}_{\mathbb{G}}^{-1}\frac{1}{\tilde{w}}\mathbf{1}_{[\alpha,\infty)}(\tilde{r})\hat{\gamma}, \qquad \alpha > 0. \tag{3.22}$$

PROOF. Just observe that

$$\begin{aligned}(\mathcal{F}_{\mathbb{G}}\hat{q})(\gamma) &= c_{\mathbb{G}}\frac{1}{n}\sum_{k=1}^{n}\int_{\mathbb{G}} w^*(x \ominus X_k)\gamma(x)d\mu_{\mathbb{G}}(x) \\ &= c_{\mathbb{G}}\frac{1}{n}\sum_{k=1}^{n}\gamma(X_k)\int_{\mathbb{G}} w^*(y)\gamma(y)d\mu_{\mathbb{G}}(y) \\ &= c_{\mathbb{G}}\widetilde{w^*}(\gamma)\hat{\gamma}(\gamma).\end{aligned} \tag{3.23}$$

Application of (3.15), where r is defined in (1.5), yields (3.22).

THEOREM 3.2. *The RI-type indirect regression estimator can be expressed as*

$$\hat{\theta}_\alpha = c_{\mathbb{G}}\mathcal{F}_{\mathbb{G}}^{-1}\frac{1}{\tilde{w}}\mathbf{1}_{[\alpha,\infty)}(\tilde{r})\hat{\chi}, \qquad \alpha > 0, \tag{3.24}$$

PROOF. A similar calculation yields here

$$(\mathcal{F}_{\mathbb{G}}\hat{q})(\gamma) = c_{\mathbb{G}}\widetilde{w^*}(\gamma)\hat{\chi}(\gamma), \tag{3.25}$$

which entails (3.24).

Let us now apply Theorem 2.3 to show that the risk of these estimators attains the optimal rate. In the examples of Section 4, some optimal rates will be explicitly determined. We choose the submodel to be of type (2.4), i.e.

$$\Theta_0 := \{\theta \in L^2(\mu_{\mathbb{G}}) : |(\mathcal{F}_{\mathbb{G}}\theta)(\gamma)| \le \ell(\gamma), \gamma \in \Gamma\}, \qquad \ell \in L^2(\mu_\Gamma). \tag{3.26}$$

Results for hyperellipses like $\{\theta \in L^2(\mu_{\mathbb{G}}) : \|\mathcal{F}_{\mathbb{G}}\theta\|_\Gamma \le c\}$, for some $0 < c < \infty$, can be obtained via hyperrectangles as in (3.26). For the lower bound we need $\Theta_0 \subset \Theta$. In the indirect density estimation case this condition and (2.8) is harder to verify than in the indirect regression case, because Θ has to remain restricted to nonnegative functions with integral 1. For convolution on the real line such a construction and verification has been carried out in van Rooij and Ruymgaart (1996). We will also need that

$$\text{(3.27)} \qquad \begin{cases} \text{conditions (2.6), (2.9), (2.13) are fulfilled with} \\ \mathbb{H} = L^2(\mu_{\mathbb{G}}), \quad \mathbb{S} = \Gamma, \quad \nu = \mu_\Gamma, \ \rho = \tilde{r}, \end{cases}$$

in order to suitably relate the decay of $\mathcal{F}_{\mathbb{G}}\theta$ to the decay of $\tilde{r}$ and the Fisher information.

THEOREM 3.3. *In the indirect density estimation model let* (2.8), (3.21), *and* (3.27) *be fulfilled, and suppose that* $\Theta_0 \subset \Theta$. *Then there exist* $\alpha = \alpha(n) > 0$ *such that the MISE of the RI-estimators* $\hat{\theta}_{\alpha(n)}$ *in* (3.22) *converges to* 0 *at the minimax rate over the submodel* Θ_0, *as* $n \to \infty$.

PROOF. This follows at once from Theorem 2.3 provided that the crucial condition (2.2) is fulfilled. To verify this condition first note that $\boldsymbol{E}|(\mathcal{F}_{\mathbb{G}}\hat{q})(\gamma) - (\mathcal{F}_{\mathbb{G}}q)(\gamma)|^2 = \operatorname{Var}(\mathcal{F}_{\mathbb{G}}\hat{q})(\gamma) = n^{-1}\operatorname{Var}(\mathcal{F}_{\mathbb{G}}\hat{q}_1)(\gamma)$. It follows from (3.23) that we have

$$\text{(3.28)} \qquad \begin{aligned} \operatorname{Var}(\mathcal{F}_{\mathbb{G}}\hat{q}_1)(\gamma) &\le \boldsymbol{E}|(\mathcal{F}_{\mathbb{G}}\hat{q}_1)(\gamma)|^2 \\ &\le c_{\mathbb{G}}^2|\widetilde{w^*}(\gamma)|^2\boldsymbol{E}|\gamma(X_1)|^2 = c_{\mathbb{G}}^2\tilde{r}(\gamma), \end{aligned}$$

because γ maps into the complex unit circle.

In the regression case we usually have $\Theta = L^2(\mu_{\mathbb{G}})$, so that the condition $\Theta_0 \subset \Theta$ will be automatically satisfied. Rather than requiring (2.8) we may require that the family of densities $f_2(\bullet - \tau)$, $\tau \in \mathbb{R}$, where f_2 is the density of the measurement errors, satisfies

$$\text{(3.29)} \qquad \begin{cases} f_2 \text{ is Lebesgue-absolutely-continuous with Radon-Nikodym} \\ \text{derivative } f_2', \text{ satisfying } \mathcal{I}_2 := \int_{-\infty}^{\infty}[\{f_2'(x)\}^2/f_2(x)]dx < \infty. \end{cases}$$

We also need to extend condition (1.12) to

$$\text{(3.30)} \qquad 0 < m := \operatorname{ess}\inf_{z\in\mathbb{G}} f_1(z) \le \operatorname{ess}\sup_{z\in\mathbb{G}} f_1(z) =: M < \infty.$$

THEOREM 3.4. *In the indirect regression model let* (3.21), (3.27), (3.29), *and* (3.30) *be fulfilled. Then there exist* $\alpha = \alpha(n) > 0$ *such that the MISE of the RI-estimators* $\hat{\theta}_{\alpha(n)}$ *in* (3.24) *converges to* 0 *at the minimax rate over the submodel* Θ_0, *as* $n \to \infty$.

PROOF. First of all (3.27), (3.29), and (3.30) entail (2.8). Assumption (3.29) implies that $\sqrt{f_2(\bullet - \tau)}$ has a strong derivative in $L^2(\mathbb{R})$ at every τ, equal to

$$\text{(3.31)} \qquad \frac{d}{d\tau}\sqrt{f_2(\bullet - \tau)} = \frac{f_2'(\bullet - \tau)}{2\sqrt{f_2(\bullet - \tau)}} \in L^2(\mathbb{R}), \qquad \tau \in \mathbb{R},$$

see Hájek & Šidák (1967). We see from (1.9) and (2.7) that for $\theta \in \Theta_0$ (i.e. $t \in I_0$) the density of the bivariate X_i equals $f_t(y,z) = f_2(y - \sum_{k=1}^{\infty} t_k(w \circledast e_k)(z))f_1(z)$, $y \in \mathbb{R}$, $z \in \mathbb{G}$. Assuming also pointwise convergence of $\sum_{k=1}^{\infty} t_k w \circledast e_k$, by (3.31) the strong partial derivatives of $\sqrt{f_t}$ exist for every $t \in I_0$ and are equal to

$$\text{(3.32)} \qquad \left(\frac{\partial\sqrt{f_t}}{\partial t_k}\right)(y,z) = \frac{f_2'(y - p_t(z))}{2\sqrt{f_2(y - p_t(z))}} \cdot (w \circledast e_k)(z)\sqrt{f_1(z)},$$

where as before $p = p_t = \sum_{k=1}^{\infty} t_k(w \circledast e_k)$. It follows from (3.30) that

$$\begin{aligned}(3.33)\qquad \left\|\frac{\partial\sqrt{f_t}}{\partial t_k}\right\|^2 &= \frac{1}{4}\int_{-\infty}^{\infty}\frac{\{f_2'(x)\}^2}{f_2(x)}dx\int_{\mathbb{G}}|(w\circledast e_k)(z)|^2 f_1^2(z)\,d\mu_{\mathbb{G}}(z)\\ &\le \left(\frac{M}{2}\right)^2\mathcal{I}_2\|w\circledast e_k\|_{\mathbb{G}}^2 = \left(\frac{M}{2}\right)^2\mathcal{I}_2\|\tilde{w}.\varphi_k\|_{\Gamma}^2\\ &= \left(\frac{M}{2}\right)^2\mathcal{I}_2\left\{\frac{1}{\mu_\Gamma(S_k)}\int_{S_k}|\tilde{w}(\gamma)|^2 d\mu_\Gamma(\gamma)\right\} = C\tilde{r}_k,\end{aligned}$$

for some $0 < C < \infty$, where $\tilde{r}_k$ is the average of $\tilde{r}$ over the set $S_k \subset \Gamma$. Since the upper bound in (3.33) is independent of the choice of $t \in I_0$, condition (2.8) is clearly satisfied for numbers ρ_k that relate to the function $\rho = \tilde{r}$ representing the convolution operator in the frequency domain.

It remains to verify condition (2.2). Exploiting (3.30) and (3.25) we now arrive at

$$\begin{aligned}(3.34)\quad \operatorname{Var}(\mathcal{F}_{\mathbb{G}}\hat{q}_1)(\gamma) &\le \boldsymbol{E}|(\mathcal{F}_{\mathbb{G}}\hat{q}_1)(\gamma)|^2\\ &\le c_{\mathbb{G}}^2|\widetilde{w^*}(\gamma)|^2\boldsymbol{E}\left(\frac{Y}{f_1(Z)}\right)^2 = c_{\mathbb{G}}^2\tilde{r}(\gamma)\boldsymbol{E}\left(\frac{1}{f_1^2(Z)}\mathbb{E}(Y^2\mid Z)\right)\\ &= c_{\mathbb{G}}^2\tilde{r}(\gamma)\int_{\mathbb{G}}\frac{1}{f_1^2(z)}\left\{\int_{-\infty}^{\infty}y^2 f_2(y-(w\circledast\theta)(z))dy\right\}f_1(z)d\mu_{\mathbb{G}}(z)\\ &= c_{\mathbb{G}}^2\tilde{r}(\gamma)\left\{\left(\frac{\sigma}{m}\right)^2+\frac{\|w\circledast\theta\|_{\mathbb{G}}^2}{m}\right\},\end{aligned}$$

so that (2.2) is satisfied indeed.

Let us now return to the problem, already discussed in Section 1, that although in principle Theorem 2.3 and its corollaries Theorem 3.3 and Theorem 3.4 determine the rate at which $\alpha(n)$ should tend to 0, as $n \to \infty$, they don't specify the choice of the smoothing parameter for any given fixed sample size. To specialize the general method in Dey *et al.* (1996) for data-driven smoothing parameter selection to abstract convolutions as in Example 1.1 let us first introduce

$$(3.35)\qquad \hat{q}_{(j)} := \frac{1}{n-1}\sum_{k\neq j}\hat{q}_k,$$

with $\hat{q}_k$ defined in (1.7). Thus we suggest to set the regularization parameter equal to

$$(3.36)\qquad \hat{\alpha} := \arg\min_{\alpha>0}\hat{M}_n(\alpha),$$

where the random function $\hat{M}_n$ is given by

$$(3.37)\quad \hat{M}_n(\alpha) := \frac{1}{n}\sum_{j=1}^{n}\int_{\{\tilde{r}\ge\alpha\}}\frac{1}{\tilde{r}^2}\left\{\frac{1}{n}|\mathcal{F}_{\mathbb{G}}\hat{q}_{(j)}|^2 - \frac{n+1}{n}(\mathcal{F}_{\mathbb{G}}\hat{q}_{(j)})\overline{(\mathcal{F}_{\mathbb{G}}\hat{q}_j)}\right\}d\mu_\Gamma,\quad \alpha > 0.$$

Note that the Fourier transforms needed for (3.37) are calculated in (3.23) or (3.25). In Dey *et al.* (1996) this selection procedure has been employed for convolution on the real line with satisfactory results.

4. Some examples

The first example concerns an errors-in-variables model involving convolution on $\mathbb{Z}^2$. It shows that deconvolution is not necessarily ill-posed and that the MISE may still converge at the rate n^{-1}. In the second example we deal with errors-in-variables involving convolution on $\mathbb{R}^2$, where the rate of the MISE depends on the smoothness class considered.

Example 4.1. Let us consider the model of Example 1.1 with $\mathbb{G} = \mathbb{Z}^2$ and density $w(m,n) := \psi(m) \cdot \psi(n)$, $(m,n) \in \mathbb{Z}^2$, where ψ is the density of the Poisson (1)-distribution, i.e.

$$(4.1)\quad \psi(k) := e^{-1}\frac{1}{k!} \quad \text{for} \quad k = 0,1,\ldots, \quad \text{and} \quad \psi(k) = 0 \;\text{ for } k = -1,-2,\ldots.$$

Then $r(m,n) = (\psi^* \circledast \psi)(m).(\psi^* * \psi)(n)$, where $\psi^*(k) := \psi(-k)$, $k \in \mathbb{Z}$. It follows that

$$(4.2)\qquad \tilde{r}(s,t) = e^{-4}e^{2(\cos s+\cos t)}, \qquad (s,t) \in [0,2\pi)^2.$$

Because apparently

$$(4.3)\qquad \min_{(s,t)\in[0,2\pi)^2} \tilde{r}(s,t) = e^{-8} > 0,$$

this deconvolution problem is not really ill-posed.

According to (1.7) the estimator of q will now be

$$(4.4)\quad \begin{aligned} \hat{q}(m,n) &= \frac{1}{n}\sum_{k=1}^{n} \psi^*(m - X_{k1})\psi^*(n - X_{k2}) \\ &= \frac{1}{n}\sum_{k=1}^{n} \psi(X_{k1} - m)\psi(X_{k2} - n) \\ &= \frac{1}{n}\left(\frac{1}{e}\right)^2 \sum_{k:X_{k1}\geq m, X_{k2}\geq n} \frac{1}{(X_{k1}-m)!(X_{k2}-n)!}, \qquad (m,n) \in \mathbb{Z}^2. \end{aligned}$$

It is clear from (4.3) that choosing $0 < \alpha < e^{-8}$ yields $R_\alpha^{-1} = R^{-1}$ and a zero bias term. For such an α the RI-estimator equals

$$(4.5)\qquad \hat{\theta}(m,n) = \sum_{k\in\mathbb{Z}}\sum_{\ell\in\mathbb{Z}} \Delta(m-k, n-\ell)\hat{q}(k,\ell), \qquad (m,n) \in \mathbb{Z}^2,$$

where Δ is given by

$$(4.6)\quad \begin{aligned} \Delta(m,n) &= \int_0^{2\pi}\int_0^{2\pi} \frac{1}{\tilde{r}(s,t)} e^{-i(ms+nt)}dsdt \\ &= e^4\left\{\int_0^{2\pi} e^{-ims-2\cos s}ds\right\}\left\{\int_0^{2\pi} e^{-int-2\cos t}dt\right\}, \qquad (m,n) \in \mathbb{Z}^2. \end{aligned}$$

This estimator has $\boldsymbol{E}\|\hat{\theta} - \theta\|^2 = \mathcal{O}(n^{-1})$, as $n \to \infty$, which rate is of course optimal.

Example 4.2. Let us now consider a special case of Example 1.1 where more typical rates arise. Suppose the bivariate density w is such that its Fourier transform satisfies

$$|\tilde{w}(s,t)| \sim C(s^2+t^2)^{-a}, \qquad \text{as} \quad s^2+t^2 \to \infty, \tag{4.7}$$

for some $0 < C < \infty$ and $a > \frac{1}{2}$. From (1.7) we see that the estimator of q is given by

$$\begin{aligned} \hat{q}(x,y) &= \frac{1}{n}\sum_{k=1}^{n} w^*(x - X_{k1}, y - X_{k2}) \\ &= \frac{1}{n}\sum_{k=1}^{n} w(X_{k1} - x, X_{k2} - y), \qquad (x,y) \in \mathbb{R}^2, \end{aligned} \tag{4.8}$$

which yields the *RI*-estimator

$$\hat{\theta}_\alpha(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Delta_\alpha(x-u, y-v)\hat{q}(u,v)dudv, \qquad (x,y) \in \mathbb{R}^2, \quad \alpha > 0, \tag{4.9}$$

where Δ_α is given by

$$\Delta_\alpha(x,y) = \frac{1}{2\pi}\iint_{\{\tilde{r}\geq\alpha\}} e^{-i(xs+yt)}\frac{1}{\tilde{r}(s,t)}dsdt, \qquad (x,y) \in \mathbb{R}^2. \tag{4.10}$$

Let us next restrict θ to the submodel

$$\{\theta \in \Theta \subset L^2(\mathbb{R}^2) : |\tilde{\theta}(s,t)| \leq C(1+s^2+t^2)^{-\nu}, (s,t) \in \mathbb{R}^2\}, \tag{4.11}$$

where $0 < C < \infty$ (we use C here as a generic constant) and $\nu > \frac{1}{2}$. This is a common way to describe smoothness of functions on $\mathbb{R}^2$, related to the Sobolev norm. According to Theorem 2.1 the MISE of the estimator $\hat{\theta}_\alpha$ satisfies

$$\begin{aligned} \boldsymbol{E}\|\hat{\theta}_\alpha - \theta\|^2 &\leq \frac{C}{n}\iint_{\{(s^2+t^2)^{-2a}\geq\alpha\}} (s^2+t^2)^{2a}dsdt \\ &\quad + C\iint_{\{(s^2+t^2)^{-2a}<\alpha\}} \frac{1}{(s^2+t^2)^{2\nu}}dsdt \\ &\leq \frac{C}{n}\int_0^A u^{4a+1}du + C\int_A^\infty u^{1-4\nu}du, \end{aligned} \tag{4.12}$$

where $A := \alpha^{-1/4a}$, as we see from using polar coordinates.

Setting $A = n^\delta$, $\delta > 0$, the two terms in the upper bound in (4.12) are of the same order if $\delta = 1/4(a+\nu)$, which yields

$$\boldsymbol{E}\|\hat{\theta}_{\alpha(n)} - \theta\|^2 = \mathcal{O}(n^{-(2\nu-1)/2(a+\nu)}), \qquad \text{as} \quad n \to \infty, \tag{4.13}$$

when we choose $\alpha(n) \sim n^{-a/(a+\nu)}$. If the proper conditions are fulfilled we may argue along the lines of Theorem 2.3 or Theorem 3.3 that the rate in (4.13) is also optimal. Let us just observe that (4.7) and (4.11) entail that (2.13) is fulfilled for

$$\Psi(\rho) = C\rho^{\nu/2a}, \qquad \rho \geq 0. \tag{4.14}$$

In conclusion let us briefly comment on (4.7). If we would choose $w(x,y) = \varphi(x).\varphi(y)$, $(x,y) \in \mathbb{R}^2$, where φ denotes the standard normal density, we would have $|\tilde{w}(s,t)| = \exp(-\frac{1}{2}(s^2+t^2))$, $(s,t) \in \mathbb{R}^2$. This function has the desired radial symmetry, but its decay is much faster than that considered in (4.7) and a much slower, logarithmic, rate would result in (4.13).

Acknowledgements

The authors are grateful to Professor Michael Nussbaum for the invitation to present part of this material at the 21st European Meeting of Statisticians, held in Aarhus, 1995. They are also grateful to the associate editor and a referee for some useful comments and very careful reading of the manuscript.

References

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates, *Biometrika*, **71**, 353–360.

Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise, *Ann. Statist.*, **24**, 2384–2398.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density, *J. Amer. Statist. Assoc.*, **83**, 1184–1186.

Carroll, R. J., van Rooij, A. C. M. and Ruymgaart, F. H. (1991). Theoretical aspects of ill-posed problems in statistics, *Acta Appl. Math.*, **24**, 113–140.

Chandrawansa, K., van Rooij, A. C. M. and Ruymgaart, F. H. (1999). Speed of convergence in the Hausdorff metric for estimators of irregular mixing densities, *J. Nonparametr. Statist.*, **10**, 375–387.

Chauveau, D., van Rooij, A. C. M. and Ruymgaart, F. H. (1994). Regularized inversion of noisy Laplace transforms, *Adv. in Appl. Math.*, **15**, 186–201.

Dey, A. K., Mair, B. A. and Ruymgaart, F. H. (1996). Cross-validation for parameter selection in inverse estimation problems, *Scand. J. Statist.*, **23**, 609–620.

Dey, A. K., Martin, C. F., and Ruymgaart, F. H. (1998). Input recovery from noisy output data, using regularized inversion of the Laplace transform, *IEEE Trans. Inform. Theory*, **44**, 1125–1130.

Diaconis, P. (1988). *Group Representations in Probability and Statistics*, Lecture Notes-Monograph Series #11, Institute of Mathematical Statistics, Hayward, California.

Donoho, D. L. (1994). Statistical estimation and optimal recovery, *Ann. Statist.*, **22**, 238–270.

Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comput. Harmon. Anal.*, **2**, 101–126.

Efromovich, S. and Pinsker, M. S. (1984). An adaptive algorithm of nonparametric filtering, *Automat. Remote Control*, **11**, 1434–1440.

Fan, J. (1991). Global behavior of deconvolution kernel estimates, *Statist. Sinica*, **1**, 541–551.

Gill, R. D. and Levit, B. Y. (1995). Applications of the van Trees inequality: A Bayesian Cramér-Rao bound, *Bernoulli*, **1**, 59–79.

Gilliam, D. S., Schulenberger, J. R. and Lund, J. H. (1988). Spectral representation of the Laplace and Stieltjes transforms, *Mat. Apl. Comput.*, **7**, 101–107.

Goldenshluger, A. (1997). On pointwise adaptive nonparametric deconvolution, Tech. Report, University of Haifa, Haifa, Israel.

Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression, *Ann. Statist.*, **18**, 758–778.

Groeneboom, P. and Jongbloed, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem, *Ann. Statist.*, **23**, 1518–1542.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

Hall, P. (1990). Optimal convergence rates in signal recovery, *Ann. Probab.*, **18**, 887–900.

Halmos, P. R. (1963). What does the spectral theorem say? *Amer. Math. Monthly*, **70**, 241–247.

Healy, D. M. Jr. and Kim, P. T. (1993). Spherical deconvolution with application to geometric quality assurance, Tech. Report, Dept. Mathematics and Computer Science, Dartmouth College.

Healy, D. M. Jr. and Kim, P. T. (1996). An empirical Bayes approach to directional data and efficient computation on the sphere, *Ann. Statist.*, **24**, 232–254.

Healy, D. M. Jr., Hendriks, H. and Kim, P. T. (1998). Spherical deconvolution, *J. Multivariate Anal.*, **67**, 1–22.

Hewitt, E. and Ross, K.A. (1963). *Abstract Harmonic Analysis*, *I*, Springer, New York.

Hoerl, A. and Kennard, R. W. (1970*a*). Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.

Hoerl, A. and Kennard, R. W. (1970*b*). Ridge regression: Application to nonorthogonal problems, *Technometrics*, **12**, 69–82.

Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems, *Ann. Statist.*, **18**, 251–280.

Kim, P. T. (1998). Deconvolution density estimation on $SO(N)$, *Ann. Statist.*, **26**, 1083–1102.

Koo, J.-Y. (1993). Optimal rates of convergence for nonparametric statistical inverse problems, *Ann. Statist.*, **21**, 590–599.

Kress, R. (1989). *Linear Integral Equations*, Springer, New York.

Mair, B. A. and Ruymgaart, F. H. (1995). Statistical inverse estimation in Hilbert scales, *SIAM J. Appl. Math.*, **56**, 1424–1444.

Mardia, K. V. and Jupp, P. (2000). *Directional Statistics*, Wiley, New York.

Margenau, H. and Murphy, G. M. (1956). *The Mathematics of Physics and Chemistry*, 2nd ed., Van Nostrand, Princeton, New Jersey.

Marron, J. S. and Tsybakov, A. B. (1995). Visual error criteria for qualitative smoothing, *J. Amer. Statist. Assoc.*, **90**, 499–507.

Neumann, M. H. (1995). Optimal change-point estimation in inverse problems, Preprint #163, Weierstrass-Institut, Berlin.

Nussbaum, M. (1996). The Pinsker bound: A review, Preprint #281, Weierstrass Institut, Berlin.

Nychka, D. and Cox, D. D. (1989). Convergence rates for regularized solutions of integral equations from discrete noisy data, *Ann. Statist.*, **17**, 556–572.

Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise, *Problems Inform. Transmission*, **16**, 120–133.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators, *Scand. J. Statist.*, **9**, 65–78.

Terras, A. (1985). *Harmonic Analysis on Symmetric Spaces and Applications*, *I*, Springer, New York.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of Ill-posed Problems*, Wiley, New York.

van Rooij, A. C. M. and Ruymgaart, F. H. (1991). Regularized deconvolution on the circle and the sphere, *Nonparametric Functional Estimation and Related Topics* (ed. G. G. Roussas), Kluwer, Dordrecht.

van Rooij, A. C. M. and Ruymgaart, F. H. (1996). Asymptotic minimax rates for abstract linear estimators, *J. Statist. Plann. Inference*, **53**, 389–402.

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*, Springer, New York.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions, *Ann. Statist.*, **18**, 806–831.