Journal of Applied Mathematics and Stochastic Analysis, **13**:1 (2000), 1-14.

CONTROLLING THE GIBBS PHENOMENON IN NOISY DECONVOLUTION OF IRREGULAR MULTIVARIABLE INPUT SIGNALS[1]

KUMARI CHANDRAWANSA and FRITS H. RUYMGAART
Texas Tech University
Department of Mathematics and Statistics
Lubbock, TX 79409 USA

ARNOUD C.M. VAN ROOIJ
Katholieke Universiteit Nijmegen
Department of Mathematics
6525 ED Nijmegen, The Netherlands

(Received January, 1998; Revised November, 1998)

An example of inverse estimation of irregular multivariable signals is provided by picture restoration. Pictures typically have sharp edges and therefore will be modeled by functions with discontinuities, and they could be blurred by motion. Mathematically, this means that we actually observe the convolution of the irregular function representing the picture with a spread function. Since these observations will contain measurement errors, statistical aspects will be pertinent. Traditional recovery is corrupted by the Gibbs phenomenon (i.e., overshooting) near the edges, just as in the case of direct approximations. In order to eliminate these undesirable effects, we introduce an integral Cesàro mean in the inversion procedure, leading to multivariable Fejér kernels. Integral metrics are not sufficiently sensitive to properly assess the quality of the resulting estimators. Therefore, the performance of the estimators is studied in the Hausdorff metric, and a speed of convergence of the Hausdorff distance between the graph of the input signal and its estimator is obtained.

Key words: Deconvolution, Irregular Multivariable Input Signals, Gibbs Phenomenon, Overshooting, Cesàro Integral Means, Hausdorff Metric, Speed of Convergence.

AMS subject classifications: 62G07, 94A12.

1. Introduction

A practical example of noisy deconvolution of irregular multivariable signals is

[1]Work partially supported by NSF grant DMS-95-04485.

provided by image reconstruction. An image may have sharp edges, and can generally be represented by an irregular function of two variables. Due to motion, for instance, the actual observed picture is a convolution of the image with a known spread function, where this convolution is observed with random measurement error. Adopting a random design in order to remain in the framework of independent and identically distributed (IID) random variables, a more precise description of the above situation is that we observe n independent copies of $X:=(Y,Z)$, where:

$$Y:=(w*\theta)(Z)+E,\ \theta\in\Theta, \tag{1.1}$$

Z is a random vector in $\mathbb{R}^d$ (the design) with $d\in\mathbb{N}$ arbitrary, E a random variable in $\mathbb{R}$ (the measurement error), w a known convolution kernel on $\mathbb{R}^d$, and θ the input signal contained in a class Θ of irregular functions on $\mathbb{R}^d$. More specifically, a certain type of discontinuity is permitted (see Section 2 for details).

Example: For easy calculation, the functions in this and subsequent examples will have two variables only, and will be of product type. More general functions could be considered but explicit calculations would be much harder in general. Suppose the input signal

$$\theta(x,y)=1_{[-1,1]}(x)1_{[-1,1]}(y),\quad (x,y)\in\mathbb{R}^2, \tag{1.2}$$

and that it is corrupted by the function

$$w(x,y)=\varphi(x)\varphi(y),\ \ (x,y)\in\mathbb{R}^2, \tag{1.3}$$

where

$$\varphi(x)=(1_{[-\frac{1}{2},\frac{1}{2}]}*1_{[-\frac{1}{2},\frac{1}{2}]})(x)\cos x,\ \ x\in\mathbb{R}. \tag{1.4}$$

The function φ is symmetric about 0 and continuously decreases from 1 at 0 to 0 at 1 and stays 0 on $[1,\infty)$. We have

$$(w*\theta)(x,y)=(\varphi*1_{[-1,1]})(x)(\varphi*1_{[-1,1]})(y),\ \ (x,y)\in\mathbb{R}^2. \tag{1.5}$$

This function has support in $[-2,2]\times[-2,2]$, and for the density of $Z:=(U,V)$ we might take the uniform density on this square. Note that this density remains bounded away from 0 on its support.

It is well-known that the sharp edges in the image cause degradation effects like ringing artifacts (Biemond, et al. [1]). Mathematically, the irregularities of θ induce overshooting in the recovered image related to the Gibbs phenomenon. In this note, we propose an improved deconvolution procedure to control the Gibbs phenomenon when recovering irregular multivariable signals from the random data $X_1,\ldots,X_n$, and we assess the quality of the proposed signal estimators in a suitable metric to be described below.

Traditionally, deconvolution involves Fourier methods. In ordinary direct approximation, the Fourier series displays the Gibbs phenomenon in the neighborhood of the edges. By Cesàro summation this undesirable effect can be remedied (Zygmund [15], Shilov [13], Walter [14]). In this paper we propose a modification of the deconvolution procedure in the same vein as Cesàro summation. This procedure is an extension of the one proposed in Chandrawansa, et al. [2] for univariable signals.

Donoho [3] and Neumann [10] also present interesting alternative approaches based on expansions in wavelet bases.

The use of the integrated mean square error doesn't seem to be appropriate when the signal to be recovered is irregular. A better insight into the quality of the estimators could be gained by using the supremum norm. However, even though the proposed estimation procedure keeps the Gibbs phenomenon under control, the estimators are still continuous functions and cannot possibly be close to the irregular input signal in the supremum norm. Therefore, the distance derived from the supremum norm is replaced by the Hausdorff distance between the extended closed graphs of the functions involved. This Hausdorff distance is related tot he sup-norm distance, and equivalent to it when the functions are continuous. Marron and Tsybakov [9] advocate use of mathematical error criteria that more closely follow visual impression, and mention the Hausdorff distance as a possibility. Korostelev and Tsybakov [8] employ the Hausdorff metric when the support of a density is estimated.

As we see from Equation (1.1), the problem considered here can be cast in the format of an indirect nonparametric function estimation problem, where this function is multivariable.

Let us conclude this introduction with an outline of our approach. The usual direct regression estimators could be used to estimate p, where

$$p = w*\theta. \tag{1.6}$$

Such estimators would require the choice of a kernel and bandwidth that are extraneous to our problem and, moreover, the resulting estimator cannot be unbiased. For this reason, and partly because the convolution operator in Equation (1.6) is not in general Hermitian (symmetry of w about 0 is required for this operator to be Hermitian), we follow a somewhat different route by replacing Equation (1.6) with an equivalent one which is easier to deal with. For this purpose we apply an injective smoothing operator to either side, a process referred to as preconditioning. One of the effects of preconditioning is that p is replaced with a smoother function q, which can be unbiasedly and $\sqrt{n}$-consistently estimated in a direct and simple manner. The effect of preconditioning can be compared to replacing a density (which cannot be unbiasedly or $\sqrt{n}$-consistently estimated) by its smoother cumulative distribution function (that does allow a simple unbiased and $\sqrt{n}$-consistent estimator).

Here we will precondition by means of convolution with $w^*(x) := w(-x)$, $x \in \mathbb{R}^d$, which yields the equation

$$q = R\theta, \text{ where } R \text{ is convolution with } r := w^* * w. \tag{1.7}$$

In order to ensure that θ is identifiable, we need to assume convolution with w to be injective, and under this assumption Equations (1.6) and (1.7) are equivalent. We will see in Section 3 that q can be unbiasedly estimated using an intrinsic method; moreover, R is a Hermitian operator which is somewhat more convenient to deal with.

Equation (1.7) is ill-posed because the inverse of the convolution is unbounded. This ill-posedness is a serious problem because we have only imperfect knowledge of q as represented in its estimator $\widehat{q}$. The inverse will be regularized by using the spectral cut-off scheme in the frequency domain. When we transform back to the "time" domain, the inverse Fourier transform will be replaced by an integral Cesàro mean, resulting in a regularized inverse $\bar{R}_M^{-1}$ of the operator R in Equation (1.7), for a suitable truncation point $M > 0$. We propose to estimate θ by

$$\widehat{\theta}_M := \bar{R}_M^{-1}\widehat{q}. \tag{1.8}$$

The unbiasedness of $\widehat{q}$ entails

$$\mathbb{E}\widehat{\theta}_M = \bar{R}_M^{-1}\mathbb{E}\widehat{q} = \bar{R}_M^{-1}q =: \theta_M. \tag{1.9}$$

Denoting the Hausdorff distance by $d_{\mathcal{H}}$, we have

$$d_{\mathcal{H}}(\widehat{\theta}_M, \theta) \leq d_{\mathcal{H}}(\widehat{\theta}_M, \theta_M) + d_{\mathcal{H}}(\theta_M, \theta). \tag{1.10}$$

Easy calculation shows that the nonrandom function θ_M is independent of r, and is in fact a *direct* approximation of θ based on a smooth Cesàro average. The asymptotics of the bias part $d_{\mathcal{H}}(\theta_M, \theta)$ for large M show that the Gibbs phenomenon has been eliminated. We briefly summarize this purely analytical result in Section 2; the details can be found in van Rooij and Ruymgaart [11]. Specification of the model Θ and some further notation is also contained in Section 2. The construction of estimators for q and θ is considered in Section 3, and Section 4 is devoted to deriving a speed of almost sure convergence to zero of the Hausdorff distance on the left-hand side of Equation (1.10). It will become clear from the analysis that no extra ill-posedness is introduced by the preconditioning procedure.

2. Some Notation and Further Preliminaries

The Fourier transform $\mathfrak{F}: L^1(\mathbb{R}^d) \rightarrow L^\infty(\mathbb{R}^d)$ is given by

$$(\mathfrak{F}f)(t) := \left(\frac{1}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} e^{i\langle t, x\rangle} f(x)dx, t \in \mathbb{R}^d, \tag{2.1}$$

where it is convenient to have the special notation

$$\widetilde{f} := (2\pi)^{d/2}\mathfrak{F}f, \quad f \in L^1(\mathbb{R}^d), \tag{2.2}$$

for the characteristic function. For any $f: \mathbb{R}^d \rightarrow \mathbb{R}$, we write

$$f^+(x) := f(x) \cdot \mathbb{1}_{[0,\infty)}(f(x)), \; x \in \mathbb{R}^d \tag{2.3}$$

for the positive part. As usual we write

$$\operatorname{sinc}\xi := \frac{\sin \xi}{\xi}, \quad \xi \neq 0, \quad \operatorname{sinc} 0 := 1, \tag{2.4}$$

and introduce

$$S(t) = \left(\frac{1}{2\pi}\right)^d \prod_{k=1}^{d} \operatorname{sinc}^2 t_k, \quad t := (t_1, \ldots, t_d) \in \mathbb{R}^d, \tag{2.5}$$

for brevity.

The extended closed graph of any bounded function $f: \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\Gamma_f := (cl\,\Gamma_{f,\ell}) \cap (cl\Gamma_{f,u}), \tag{2.6}$$

where $\Gamma_{f,\ell} := \{(x,\xi) \in \mathbb{R}^{d+1}: x \in \mathbb{R}^d, \xi \leq f(x)\}$ and $\Gamma_{f,u} := \{(x,\xi) \in \mathbb{R}^{d+1}: x \in \mathbb{R}^d, \xi \geq f(x)\}$. The Hausdorff distance between any two such functions is defined as

$$\begin{aligned} d_{\mathcal{H}}(f,g) := \inf_{\varepsilon > 0}\{&\forall p \in \Gamma_f \exists q \in \Gamma_g: \| p - q \| < \varepsilon, \\ &\forall q \in \Gamma_g \exists p \in \Gamma_f: \| q - p \| < \varepsilon\}, \end{aligned} \tag{2.7}$$

where $\| \cdot \|$ is the norm in any Euclidean space ($\mathbb{R}^{d+1}$ in this case). If a function $f: \mathbb{R}^d \to \mathbb{R}$ satisfies

$$| f(x) - f(y) | \leq c \| x - y \| \ \forall x, y \in A \subseteq \mathbb{R}^d, \text{ for some } 0 < c < \infty, \tag{2.8}$$

it is called Lipschitz on A. Let us now proceed and describe the class of functions in which the model Θ is contained.

Definition: A function $f: \mathbb{R}^d \to \mathbb{R}$ is an element of the class $\mathfrak{D}$ if and only if it is bounded and integrable, and if there exists a closed, nondegenerate hypertriangle $\Delta = \Delta(f)$ such that every $a \in \mathbb{R}^d$ is contained in a hyperrectangle Δ_a, congruent to Δ, on which f is Lipschitz. The constant c appearing in the Lipschitz condition does not depend on the point $a \in \mathbb{R}^d$.

We have aimed at greatest generality in the definition of a suitable class $\mathfrak{D}$. Now we will give an explicit construction of a family of functions that is contained in $\mathfrak{D}$. Let B_ρ be the open ball in $\mathbb{R}^d$, with center at the origin and radius $\rho > 0$, and let $\bar{B}_\rho$ denote the closed ball. Suppose that $\Phi_k: B_\rho \to \mathbb{R}^d$ is a diffeomorphism for some $\rho > 1$, and let

$$A_k := \Phi_k(\bar{B}_1), \quad k = 1, \ldots, K. \tag{2.9}$$

The following result is proved in van Rooij and Ruymgaart [11].

Theorem 1: *Let the Φ_k be such that the A_k are pairwise disjoint. If f is bounded and integrable on $\mathbb{R}^d$, and Lipschitz on each A_k and on $\bigcap_{k=1}^K (A_k)^c$, then f is in $\mathfrak{D}$. In particular, we have*

$$f := \sum_{k=1}^{K} c_k \mathbb{1}_{A_k} \in \mathfrak{D}, \ c_1, \ldots, c_K \in \mathbb{R}. \tag{2.10}$$

As a special case we may take the A_k to be bounded convex sets with nonempty interior.

Rather than the usual inverse Fourier transform $\mathfrak{F}^{-1}$, we will employ $\bar{\mathfrak{F}}_M^{-1}$, defined by

$$(\bar{\mathfrak{F}}_M^{-1} g)(x) := \left(\frac{1}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} g(t) \prod_{k=1}^{d} \left(1 - \frac{| t_k |}{M}\right)^+ e^{-i\langle x, t\rangle} dt, M > 0, \tag{2.11}$$

for any $g \in L^\infty(\mathbb{R}^d)$. When we substitute $g := \mathfrak{F} f$, $f \in L^1(\mathbb{R}^d)$, we obtain

$$(\bar{\mathfrak{F}}_M^{-1}\mathfrak{F}f)(x) = \int_{\mathbb{R}^d} f\left(x - \frac{t}{M}\right)S(t)dt, \quad x \in \mathbb{R}^d, \tag{2.12}$$

where S is defined in Equation (2.5). For a proof of the result below, see van Rooij and Ruymgaart [11].

Theorem 2: *For each* $f \in \mathfrak{D}$ *there exists* $0 < c: = c(f) < \infty$ *such that*

$$d_{\mathcal{H}}(\bar{\mathfrak{F}}_M^{-1}\mathfrak{F}f, f) \leq \frac{c}{\sqrt{M}}. \tag{2.13}$$

for all $0 < M < \infty$.

3. Construction and Elementary Properties of the Estimators

First, we formulate the basic assumption from which estimators can be constructed. The conditions on the convolution kernel w partly serve to ensure the identifiability of the unknown parameter. The condition on the density of the design vector entails that an unbiased estimator of q can be constructed. Since we are free to choose this density, the condition can always be satisfied. In the next section, a more restrictive assumption will be needed to obtain a speed of convergence for the Hausdorff distance.

Basic assumptions: The model satisfies

$$\Theta \subset \mathfrak{D}, \tag{3.1}$$

and for the convolution kernel we have

$$w \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d), \text{ with } |\tilde{w}| > 0 \text{ on } \mathbb{R}^d. \tag{3.2}$$

Let the probability density f of Z concentrate mass 1 on a convex set $\sum \subset \mathbb{R}^d$ such that

$$\sum \supset \mathrm{cl}\left(\bigcup_{\theta \in \Theta}\{x \in \mathbb{R}^d: (w*\theta)(x) > 0\}\right), \tag{3.3}$$

and let f be continuous and strictly positive on $\sum$. The design vector Z is stochastically independent of the error variable E; the latter variable has a finite first moment and mean 0.

Choosing the density f of Z in accordance with this assumption allows us to construct an unbiased estimator of q, namely

$$\hat{q}(x): = \frac{1}{n}\sum_{k=1}^{n} \frac{Y_k w^*(x - Z_k)}{f(Z_k)}, \quad x \in \mathbb{R}^d. \tag{3.4}$$

To verify the unbiasedness of this estimator, we first observe that

$$\mathbb{E}(Y \mid Z) = (w*\theta)(Z). \tag{3.5}$$

Existence and evaluation of this conditional expectation follows easily from the assumption above, which entails

$$\mathbb{E}\left(\frac{Yw^*(x-Z)}{f(Z)}\right) = \mathbb{E}\mathbb{E}\left(\frac{Yw^*(x-Z)}{f(Z)} \mid Z\right)$$

$$= \mathbb{E}\left(\frac{w^*(x-Z)}{f(Z)}\mathbb{E}(Y \mid Z)\right) = \mathbb{E}\,\frac{w^*(x-Z)(w*\theta)(Z)}{f(Z)}$$

$$= \int_{\Sigma} w^*(x-z)(w*\theta)(z)dz = (w^* * w*\theta)(x) \tag{3.6}$$

$$= (r*\theta)(x) = q(x).$$

The condition in Equation (3.3) which ensures that the support of $w*\theta$ is contained in the density f, is imperative to obtain unbiasedness. Note that we also have

$$\widehat{q} \in L^1(\mathbb{R}^d) \cap L^{\infty}(\mathbb{R}^d), \text{ with } \mathbb{E}\int_{\mathbb{R}^d} \| \widehat{q}(x) \| \, dx < \infty. \tag{3.7}$$

It follows immediately from Equation (3.2) that $|Y| \leq c + |E|$ for some $0 < c < \infty$, so that

$$\begin{aligned}\mathbb{E}\left\{\frac{Y}{f(Z)}\right\}^p &\leq \mathbb{E}(c + |E|)^p \cdot \mathbb{E}\left(\frac{1}{f(Z)}\right)^p \\ &= \mathbb{E}(c + |E|)^p \cdot \int_{\Sigma}\left\{\frac{1}{f(z)}\right\}^{p-1} dz,\end{aligned} \tag{3.8}$$

for $p > 0$. Even if we require $\mathbb{E}\,|E|^p < \infty$, the last expression on the right-hand side of Equation (3.8) is still ∞ if $\sum = \mathbb{R}^d$ and $p > 1$, because the integral is inevitably ∞. In other words, if the support of $w*\theta$ is all of $\mathbb{R}^d$, the estimator $\widehat{q}(x)$ does not have any moment of order $p > 1$. For a more precise analysis of the error, we cannot avoid more restrictive conditions, as we have already noted before.

Let us next observe that $\mathfrak{F}\widehat{q}$ is well defined by Equation (3.7) and equals

$$\begin{aligned}(\mathfrak{F}\widehat{q})(t) &= \frac{1}{n}\sum_{k=1}^{n}\frac{Y_k}{f(Z_k)}(\mathfrak{F}w^*(\bullet - Z_k))(t) \\ &= (\mathfrak{F}w^*)(t)\widehat{\rho}(t), \quad t \in \mathbb{R}^d,\end{aligned} \tag{3.9}$$

where

$$\widehat{\rho}(t) := \frac{1}{n}\sum_{k=1}^{n}\frac{Y_k}{f(Z_k)}e^{i\langle Z_k, t\rangle}, \quad t \in \mathbb{R}^d. \tag{3.10}$$

Exploiting Equation (3.5) once more, it follows immediately that

$$\rho(t) := \mathbb{E}\widehat{\rho}(t) = \widetilde{w}(t)\widetilde{\theta}(t), \quad t \in \mathbb{R}^d, \tag{3.11}$$

so that $\mathfrak{F}\widehat{q}$ and $\widehat{\rho}$ are unbiased estimators of $\mathfrak{F}q$ and ρ, respectively.

Finally, let us introduce the estimator of θ and first specify the regularized inverse $\bar{R}_M^{-1}$ of R (cf. Equations (1.8) and (1.9)). With $\bar{\mathfrak{F}}_M^{-1}$ as in Equation (2.11), we define

$$\bar{R}_M^{-1}f: = \bar{\mathfrak{F}}_M^{-1}\frac{1}{\widetilde{r}}\mathfrak{F}f, \quad f \in L^1(\mathbb{R}^d), \text{ where } \widetilde{r} = |\widetilde{w}|^2. \tag{3.12}$$

This suggests the following estimator for θ:

$$\widehat{\theta}_M: = \bar{R}_M^{-1}\widehat{q}, \quad \text{for suitable } M > 0. \tag{3.13}$$

We see from Equation (3.9) that

$$\widehat{\theta}_M = \bar{\mathfrak{F}}_M^{-1}\frac{\mathfrak{F}w^*}{\widetilde{r}}\widehat{\rho} = \left(\frac{1}{2\pi}\right)^{d/2}\bar{\mathfrak{F}}_M^{-1}\left(\frac{1}{\widetilde{w}}\widehat{\rho}\right), \tag{3.14}$$

which indicates that the extra ill-posedness due to the preconditioning is compensated by the properties of $\widehat{q}$.

The estimator $\widehat{\theta}_M$ is not in general unbiased (unless we restrict Θ, for example, to bandlimited signals; cf. Ruymgaart [11]). Its expectation equals

$$\theta_M: = \mathbb{E}\widehat{\theta}_M = \left(\frac{1}{2\pi}\right)^{d/2}\bar{\mathfrak{F}}_M^{-1}\frac{1}{\widetilde{w}}\widetilde{w}\widetilde{\theta} = \bar{\mathfrak{F}}_M^{-1}\mathfrak{F}\theta, \tag{3.15}$$

which apparently does not depend on w, and Equation (2.13) applies. It is obvious that

$$\widehat{\theta}_M \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) \cap C(\mathbb{R}^d), \text{ a.s., } \theta_M \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) \cap C(\mathbb{R}^d). \tag{3.16}$$

Therefore, the Hausdorff distance between $\widehat{\theta}_M$ and θ_M essentially boils down to the distance in the supremum norm.

Example: Referring to the Example in Section 1, let us first note that Equation (3.1) is fulfilled since θ is a simple indicator (Theorem 2.1). To see that Equation (3.2) is also satisfied, it suffices to prove that $|\widetilde{\varphi}| > 0$ on $\mathbb{R}$. We write, for brevity,

$$\Delta(x) = (1_{[-\frac{1}{2},\frac{1}{2}]} * 1_{[-\frac{1}{2},\frac{1}{2}]})(x), \quad x \in \mathbb{R}, \tag{3.17}$$

and observe that

$$\begin{aligned}
\widetilde{\varphi}(t) &= \int_{-\infty}^{\infty} e^{itx}\Delta(x)\cos x\,dx \\
&= \frac{1}{2}\int_{-\infty}^{\infty} e^{itx}\Delta(x)(e^{-ix} + e^{ix})dx \\
&= \frac{1}{2}\left\{\int_{-\infty}^{\infty} e^{i(t-1)x}\Delta(x)dx + \int_{-\infty}^{\infty} e^{i(t+1)x}\Delta(x)dx\right\}
\end{aligned}$$

$$= \frac{1}{2}\left\{ \widetilde{\Delta}(t-1) + \widetilde{\Delta}(t+1) \right\}$$

$$= \frac{1}{2}\left\{ \operatorname{sinc}^2\frac{1}{2}(t-1) + \operatorname{sinc}^2\frac{1}{2}(t+1) \right\} > 0, \quad t \in \mathbb{R},$$

where $\operatorname{sinc} x = (\sin x)/x$ for $x \neq 0$, and $\operatorname{sinc} 0 = 1$. Here we might take $\Theta = \{\theta\}$, so that Equation (3.3) is satisfied, as we have seen in Section 1.

Because $w^* = w$, we have $q = w*w*\theta$, and the estimator for q (cf. Equation (3.4)) will be

$$\widehat{q}(x,y) = \frac{16}{n}\sum_{k=1}^{n} Y_k \varphi(x - U_k)\varphi(y - V_k), (x,y) \in \mathbb{R}^2. \tag{3.19}$$

Note that the factor 16 is due to the fact that the design density is uniform on $[-2,2] \times [-2,2]$. It follows that (see Equation (3.9))

$$(\mathfrak{F}\widehat{q})(s,t) = \frac{16}{n}(\mathfrak{F}w)(s,t) \sum_{k=1}^{n} Y_k e^{i(sU_k + tV_k)}$$

$$= \frac{16}{2\pi n} \sum_{k=1}^{n} Y_k e^{isU_k}\ \widetilde{\varphi}(s) e^{itV_k}\ \widetilde{\varphi}(t), \quad (s,t) \in \mathbb{R}^2, \tag{3.20}$$

with $\widetilde{\varphi}$ as in Equation (3.18).

We do not need $\widehat{q}$ or $\mathfrak{F}\widehat{q}$ for the computation of the estimator $\widehat{\theta}_M$ of actual interest. In fact, we see from Equations (3.14) and (2.11) that

$$\widehat{\theta}_M(x,y) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\frac{16}{n}\sum_{k=1}^{n} Y_k e^{i(sU_k + tV_k)}}{\widetilde{\varphi}(s)\widetilde{\varphi}(t)}$$

$$\times\left(1 - \frac{|s|}{M}\right)^+ \left(1 - \frac{|t|}{M}\right)^+ ds\,dt = \frac{16}{4\pi^2 n} \sum_{k=1}^{n} Y_k \left\{ \int_{-M}^{M} \frac{e^{isU_k}}{\widetilde{\varphi}(s)}\left(1 - \frac{|s|}{M}\right) ds \right\}$$

$$\times \left\{ \int_{-M}^{M} \frac{e^{itV_k}}{\widetilde{\varphi}(t)}\left(1 - \frac{|t|}{M}\right) dt \right\}, (x,y) \in \mathbb{R}^2. \tag{3.21}$$

4. Speed of Convergence in the Hausdorff Metric

Although under the current assumptions the integrated mean square error $\mathbb{E}\,\|\widehat{\theta}_a - \theta\|^2$ is not in general well-defined (cf. Equation (3.8)), we might employ $\mathbb{E}\,\|\widehat{\theta}_a - \theta\|$ as follows from Equation (3.7). As we have argued in the introduction, however, we prefer to use the Hausdorff metric, which for our purposes is more sensitive than an integral metric. In order to obtain a speed of convergence some further assumptions are needed. Most of these assumptions seem to be specific for the regression model, and are not needed in the errors-in-variables model, where both w and θ are probability densities and the data $X_1, \ldots, X_n$ are a random sample from the density $p = w*\theta$ (see Chandrawansa, et al. [2] for $d = 1$). It is not uncommon for convolution in a regression model to assume that both the input signal and the kernel have bounded support, which is precisely the most important of the extra conditions that will be needed here. In a practical situation (e.g., the picture restoration model) this condition is naturally satisfied. We will also assume the error variable to be bounded, although this condition could be weakened at the price of a lower rate. An assumption about the degree of ill-posedness of this deconvolution problem is inevitable if a speed of convergence is to be obtained. We will assume the convolution operator to be finitely smoothing. The infinitely smoothing case could be dealt with in a similar manner.

Further assumptions: The model satisfies

$$\Theta = \Theta_s \subset \mathfrak{D}, \tag{4.1}$$

where the signals in Θ_s have support contained in a hyperrectangle $[-\sigma, \sigma]^d$, for some known $0 < \sigma < \infty$. The convolution kernel

$$w \text{ has support in } [-\sigma, \sigma]^d, \tag{4.2}$$

where σ can be taken to be the same as in Equation (4.1). We will assume the convolution operator to be finitely smoothing, which can be described in a couple of ways in this multivariate setting. Here, we will adopt the requirement that there exists $\nu_1, \ldots, \nu_d > 0$ such that

$$\liminf_{|t_1|, \ldots, |t_d| \to \infty} \prod_{k=1}^{d} |t_k|^{\nu_k} |\widetilde{w}(t)| > 0. \tag{4.3}$$

The larger the exponents, the smaller the rate will be that we can derive. Therefore, one should choose the $\nu_1, \ldots, \nu_d$ as small as possible. Finally, we assume the boundedness of the error variable, i.e.,

$$|E| \leq \beta < \infty, \tag{4.4}$$

where it is not important that the number β be known.

It follows immediately from Equations (4.1) and (4.2) that $w * \theta$ has support in $[-2\sigma, 2\sigma]^d$. For the density f we will now take the uniform density

$$f(z) := \left(\frac{1}{4\sigma}\right)^d \mathbb{1}_{[-2\sigma, 2\sigma]^d}(z), \quad z \in \mathbb{R}^d, \tag{4.5}$$

so that our estimators in Equations (3.4), (3.10) and (3.14) simplify accordingly. Because $\widehat{\theta}_M$ and θ_M are continuous, moreover, we have $d_{\mathcal{H}}(\widehat{\theta}_M, \theta_M) \leq \| \widehat{\theta}_M - \theta_M \|_\infty$, where $\| \cdot \|_\infty$ denotes the supremum norm. It follows that

$$d_{\mathcal{H}}(\widehat{\theta}_M, \theta) \leq \| \widehat{\theta}_M - \theta_M \|_\infty + d_{\mathcal{H}}(\theta_M, \theta). \tag{4.6}$$

By Equations (3.15) and (2.13), we have

$$d_{\mathcal{H}}(\theta_M, \theta) = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right), \text{ as } M \to \infty, \tag{4.7}$$

so that our main concern now will be the first term on the right-hand side in Equation (4.6). From Equations (2.12), (3.9), (3.13) and (3.14), we see that, for any $0 < M < \infty$,

$$\| \widehat{\theta}_M - \theta_M \|_\infty = \| \mathfrak{F}_M^{-1} \frac{\widehat{\rho} - \rho}{\widetilde{w}} \|_\infty$$
$$\leq \left(\frac{1}{2\pi}\right)^{d/2} \int_{t \in [-M, M]^d} \left| \frac{\widehat{\rho}(t) - \rho(t)}{\widetilde{w}(t)} \right| \prod_{k=1}^{d} \left(1 - \frac{|t_k|}{M}\right)^+ dt. \tag{4.8}$$

At this point, we need the following result. For any fixed $\gamma > 0$, we have

$$\sup_{t \in [-n^\gamma, n^\gamma]^d} |\widehat{\rho}(t) - \rho(t)| =_{a.s.} \mathcal{O}\left(\sqrt{\frac{\log n}{n}}\right), \text{ as } n \to \infty. \tag{4.9}$$

The proof is very similar to the one for univariate empirical characteristic functions in Chandrawansa et al. [2], and is deferred to the Appendix. Exploiting Equation (4.3), which entails ($\nu = \nu_1 + \ldots + \nu_d$),

$$\sup_{t \in [-n^\gamma, n^\gamma]^d} \left| \frac{\widehat{\rho}(t) - \rho(t)}{\widetilde{w}(t)} \right| =_{a.s.} \mathcal{O}\left(n^{-\frac{1}{2} + \gamma\nu} \sqrt{\log n}\right), \text{ as } n \to \infty, \tag{4.10}$$

and choosing $M := M(n) = n^\gamma$, we obtain

$$\| \widehat{\theta}_M - \theta_M \|_\infty =_{a.s} \mathcal{O}\left(n^{-\frac{1}{2} + \gamma\nu} \sqrt{\log n}\right) \cdot \left\{ \int_0^{M(n)} \left(1 - \frac{s}{M(n)}\right) ds \right\}^d$$
$$=_{a.s.} \mathcal{O}\left(n^{-\frac{1}{2} + \gamma(\nu + d)} \sqrt{\log n}\right), \text{ as } n \to \infty. \tag{4.11}$$

For the same choice of M, the distance in Equation (4.7) will be of order $\mathcal{O}(n^{-\gamma/2})$, as $n \to \infty$. The best choice of γ will make the two terms on the right-hand side of Equation (4.6) of (approximately) the same order, which leads to taking $\gamma = 1/(2\nu + 2d + 1)$. Summarizing, we have proved our main result.

Theorem 3: *Under the assumptions of Section* 3 *and Section* 4, *we have*

$$d_{\mathcal{H}}(\widehat{\theta}_{M(n)}, \theta) =_{a.s.} \mathcal{O}(n^{-1/(4\nu + 4d + 2)} \sqrt{\log n}), \text{ as } n \to \infty, \tag{4.12}$$

provided that we take $M(n) := n^{1/(2\nu + 2d + 1)}$, *and where* $\nu = \nu_1 + \ldots + \nu_d$.

Example: Returning once again to the examples of Sections 1 and 3, it is easy to see from Equations (1.3) and (3.18) that w satisfies Equations (4.2) and (4.3), with $\nu_1 = \nu_2 = 2$, respectively. Hence, the indicator in Equation (1.2) can be recovered from the blurred convolution at a rate $\mathcal{O}(n^{-1/18} \sqrt{\log n})$.

Remarks: Rates in the literature do not easily compare with ours. Most authors not only employ the L^2-norm, but also compute the rate over smoothness subclasses in L^2 (Fan [4]). The class of interest to us is clearly not a smoothness class, and as mentioned above, the L^2-norm is not sensitive to the Gibbs phenomenon. In fact, without Cesàro averaging, the bias term would not converge at all in the Hausdorff metric employed here, but it would still converge in the L^2 norm.

Results closest to ours are certain sup-norm rates for input signals that are at least continuous. In the case of estimating a one-dimensional density, whose k-th derivative satisfies a Lipschitz condition of order α, the optimal rate in the sup-norm is of order $n^{-(\alpha + k)/(2\alpha + 2k + 1)}$, apart from a logarithmic factor. For a survey of this kind of results, see Schipper [12].

Referring to our model, if $\theta \in \Theta_\sigma$ happens to be uniformly continuous, it has been shown in van Rooij and Ruymgaart [11] that

$$d_{\mathcal{H}}(\theta_M, \theta) = \mathcal{O}\left(\frac{\log M}{M}\right), \text{ as } M \to \infty. \tag{4.13}$$

Since there will be no change in the order of magnitude of $\| \widehat{\theta}_{M(n)} - \theta_{M(n)} \|_\infty$ in Equation (4.11), we now arrive at the better rate

$$d_{\mathcal{H}}(\widehat{\theta}_{M(n)}, \theta) =_{a.s.} \mathcal{O}(n^{-1/(2\nu + 2d + 2)} \log n), \text{ as } n \to \infty, \tag{4.14}$$

provided that we take $M(n) = n^{1/(2\nu + 2d + 2)}$. In order to compare with the result in the preceding paragraph, we should take $d = 1$ and formally set $\nu = 0$ so that we obtain the rate $n^{-1/4}$, apart from a logarithmic factor. This corresponds to $k = 0$ and $\alpha = 1$ above, yielding a rate $n^{-1/3}$, apart from the logarithm. We cannot explain this discrepancy except by conjecturing that Cesàro averaging may not be desirable if it is not really necessary, like in the continuous case. It might also be possible that other averaging techniques might lead to better results, particularly for the bias term.

Our estimators are one-step linear estimators. Hall [5] suggested that better rates might be obtained by first eliminating the discontinuous part (change "points" can be rather precisely estimated, see Hinkley [6]), and then estimating the remaining smooth part. Wavelet methods, as mentioned in the introduction, might be promising, but studying wavelet expansions in the Hausdorff metric might lead to difficulties.

Acknowledgements

The authors are grateful to a referee for some useful comments.

References

[1] Biemond, J., Lagendijk, R.L. and Mersereau, R.M., Iterative methods for image deblurring, *Proc. IEEE* **78** (1990), 856-883.

[2] Chandrawansa, K., van Rooij, A.C.M. and Ruymgaart, F.H., Speed of convergence in the Hausdorff metric for estimators of irregular mixing densities, *Nonpar. Statist.* (to appear).

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

[4] Fan, J., Global behavior of deconvolution kernel estimates, *Statist. Sin.* **1** (1991), 541-551.

[5] Hall, P., Personal communication 1997.

[6] Hinkley, D.V., Inference about the change-point in a sequence of random variables, *Biometrika* **57** (1970), 1-17.

[7] Hoeffding, W., Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13-30.

[8] Korostelev, A.P. and Tsybakov, A.B., *Minimax Theory of Image Reconstruction*, Springer-Verlag, New York 1993.

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

[10] Neumann, M.H., *Optimal Change-Point Estimation in Inverse Problems*, Reprint **163**, Weierstrass-Institut, Berlin 1995.

[11] van Rooij, A.C.M. and Ruymgaart, F.H., *Speed of Convergence in the Hausdorff Metric of Fourier-Cesàro Approximations to Irregular Multivariable Functions.* Tech Report **9703**, Dept. of Math, Katholieke Universiteit, Nijmegen 1997.

[12] Schipper, M., *Sharp Asymptotics in Nonparametic Estimation*, Dissertation, University of Utrecht, The Netherlands 1997.

[13] Shilov, G.E., *Elementary Functional Analysis*, The MIT Press, Cambridge, MA 1974.

[14] Walter, G.G., *Wavelets and Other Orthogonal Systems with Applications*, CRC Press, Boca Raton, FL 1994.

[15] Zygmund, A., *Trigonometric Series, I and II*, Cambridge University Press, Cambridge, UK 1990.

Appendix

In order to prove Equation (4.9), let us take $\mathfrak{T}: = \{k \cdot n^{-\frac{1}{2}}: k = -\lfloor n^{\gamma + \frac{1}{2}} \rfloor + 1, \ldots, n^{\gamma + \frac{1}{2}} \rfloor + 1\}^d$, $\mathfrak{I} = [0, n^{-\frac{1}{2}}]^d$, and consider

$$p_n: = \mathbb{P}\left\{ \sup_{t \in [-n^{\gamma}, n^{\gamma}]^d} |\widehat{\rho}(t) - \rho(t)| \geq \lambda\sqrt{\frac{\log n}{n}} \right\}, \tag{A.1}$$

for some $\lambda > 0$. We obviously have

$$\begin{aligned} p_n \leq \mathbb{P}\left\{ \max_{t \in \mathfrak{T}} |\widehat{\rho}(t) - \rho(t)| \geq \frac{\lambda}{2}\sqrt{\frac{\log n}{n}} \right\} \\ + \mathbb{P}\left\{ \max_{t \in \mathfrak{T}} \sup_{s \in \mathfrak{I}} |\widehat{\rho}(t+s) - \widehat{\rho}(t) - [\rho(t+s) - \rho(t)]| \geq \frac{\lambda}{2}\sqrt{\frac{\log n}{n}} \right\} =: p_{n1} + p_{n2}, \end{aligned} \tag{A.2}$$

and according to the Borel-Cantelli lemma it suffices to show that a sufficiently large $\lambda > 0$ exists such that $\sum_{n=1}^{\infty} p_{nj} < \infty$ for $j = 1, 2$.

It suffices to prove the lemma for the real parts of $\widehat{\rho}$ and ρ, since the proof for the imaginary parts is similar. Hence, we let

$$\widehat{\rho}(t): = \frac{(4\alpha)^d}{n}\sum_{k=1}^{n} Y_k \cos(\langle t, Z_k \rangle), \rho(t): = (4\alpha)^d \mathbb{E} Y \cos(\langle t, Z \rangle), t \in \mathbb{R}^d. \tag{A.3}$$

Combining the boundedness of w (see Equation (3.2)) with the integrability of θ (see the definition of $\mathfrak{D}$) or conversely, with Equation (3.4), it is clear that Y assumes values in a bounded interval independent of $t \in \mathbb{R}^d$. Hence Hoeffding's [7] inequality applies to p_{n1} and yields

$$p_{n1} \le an^{d(\frac{1}{2}+\gamma)} \exp(-\lambda^2 b \log n), \tag{A.4}$$

for some $a, b \in (0,\infty)$. Apparently, we have

$$\sum_{n=1}^{\infty} p_{n1} < \infty, \text{ for any } \lambda > \left\{d\left(\frac{1}{2}+\gamma\right)/b\right\}^{\frac{1}{2}}. \tag{A.5}$$

Applying the mean value theorem and exploiting the fact that Y and Z are bounded, we easily see that

$$\sup_{s \in \mathfrak{I}} |\,\widehat{\rho}(t+s) - \widehat{\rho}(t) - [\rho(t+s) - \rho(t)]\,| \le cn^{-\frac{1}{2}}, \tag{A.6}$$

for some number $c \in (0,\infty)$ which is independent of t. This trivially entails that $p_{n2} = 0$ for all $n \ge 2$, provided that we choose $\lambda > 2c/\sqrt{\log 2}$, and Equation (4.9) follows.