The Annals of Statistics
1992, Vol. 20, No. 2, 944–970

RENORMALIZATION EXPONENTS AND OPTIMAL POINTWISE RATES OF CONVERGENCE

BY DAVID L. DONOHO AND MARK G. LOW

University of California, Berkeley

Simple renormalization arguments can often be used to calculate optimal rates of convergence for estimating linear functionals from indirect measurements contaminated with white noise. This allows one to quickly identify optimal rates for certain problems of density estimation, nonparametric regression, signal recovery and tomography. Optimal kernels may also be derived from renormalization; we give examples for deconvolution and tomography.

1. Introduction. Let $f = f(t)$, $t \in R^d$, be an unknown object (signal, intensity, density, ...), a real valued function on R^d, and suppose we are interested in recovering the linear functional $T(f)$, for example $f(0)$ or $f^{(k)}(0)$, $k > 0$. We know a priori that $f \in \mathscr{F}$, a certain convex class of functions (e.g., a class of smooth functions). Depending on the type of measurements we are able to make, problems of this form arise in statistical settings, such as nonparametric density estimation and nonparametric regression estimation; but they also arise in signal recovery and remote sensing.

In such problems it has been observed that there exists an optimal rate of convergence at which the mean squared error in estimation of T from n data can go to zero as n increases. For example, suppose we are in the density estimation problem, where we observe data $X_1, \dots, X_n$ which are random samples from an unknown density f on R^1, and we wish to estimate $T(f) = f(0)$. We know a priori only that $f \in \mathscr{F}_{2,\infty} = \{f\colon \|f''\|_\infty \le 1, \|f\|_\infty \le M\}$. Define the minimax mean squared error

$$R(n) = \inf_{\hat{T}} \sup_{\mathscr{F}} E\big(\hat{T}(X_1, \dots, X_n) - T(f)\big)^2.$$

Rosenblatt (1956) showed essentially that $R(n) \le Cn^{-4/5}$; and Farrell (1972) proved that $R(n) \ge cn^{-4/5}$; thus $n^{-4/5}$ is the optimal rate at which the mean squared error of an estimate can go to zero with increasing n; see Wahba (1975), Meyer (1977), Ibragimov and Hasminskii (1981), Stone (1980) and Hall (1989) for optimal rate calculations in related density estimation problems with various choices of T and $\mathscr{F}$. In nonparametric regression estimation, we gather n observations $y_i = f(t_i) + z_i$, with (t_i) and (z_i) i.i.d. sequences; opti-

Received June 1990; revised June 1991.
[1]Supported by NSF Grants DMS-84-51753 and DMS-88-10192, by grants from Apple Computer, A.T.T. Foundation, Schlumberger-Doll Research and Western Geophysical.
AMS 1980 *subject classifications*. Primary 62G07; secondary 62C20.
Key words and phrases. Radon transform, Riesz transform, deconvolution, partial deconvolution, minimax kernels, boundary kernels, minimax linear estimation, minimax risk, white noise model, Gaussian experiments.

mal rates of convergence $n^{-r(T,\mathscr{F})}$ for estimation of functionals $T(f)$ under various smoothness classes $\mathscr{F}$ have been established by Stone (1980), Brown and Low (1991), among others. In tomography and problems of remote sensing, one observes a finite number n of noisy, indirect observations; these also give rise to optimal rates $n^{-r(T,\mathscr{F})}$ for the minimax mean squared error. Finally, Hall (1990) considers a signal recovery problem—removal of out-of-focus blur in images—and, modelling it with a sequence of approximating problems, indexed by n, finds again that the minimax mean squared error goes to zero at a definite rate $n^{-r(T,\mathscr{F})}$, depending on the function class to which the unknown object belongs and on the extent of out-of-focus blurring.

The rates phenomenon raises two immediate questions:

1. What determines the value of $r(T,\mathscr{F})$?
2. How can one efficiently calculate it?

In this paper, we show that the notion of *renormalization* provides a convenient answer to both questions. For a certain class of linear problems, we identify the existence of rates of convergence with easily measured scaling properties of the function class $\mathscr{F}$, the functional T and the observation scheme. In Section 3, we describe a simple procedure which for many problems allows one to calculate optimal rates easily and quickly. The procedure is to identify the renormalization exponents s in relations of the form

$$J(af(b\,\cdot)) = ab^s J(f(\cdot))$$

for three key homogeneous functionals J associated with the estimation problem. The rate functional is then a simple combination of the three exponents. Section 4 below shows how such arguments can be used for determining optimal rates in problems of density estimation and tomography.

A second use for renormalization is in the derivation of optimal kernels, that is, of minimax mean square linear estimators. In essence, the optimal kernel is identified as the solution of a certain extremal problem and the optimal bandwidth as the renormalization constant which obtains this extremal problem from the same extremal problem in normal form. In Section 8 we derive optimal kernels for some problems of tomography and deconvolution.

Here we develop both aspects of renormalization—determining rates and determining optimal kernels—into a smoothly functioning tool. The generality of this tool is considerably helped by the fact, shown in Sections 5 and 6, that many problems which do not renormalize exactly do renormalize in an asymptotic sense, and the simple rate derivation continues to apply in such cases. Section 7 below describes a four-step renormalization heuristic which can allow one to efficiently derive the optimal rates of convergence in a variety of problems which do not renormalize exactly, for example, in signal recovery and in partial deconvolution.

The first explicit use of renormalization in identifying optimal rates occurs in Low (1992). The first mention of renormalization for obtaining optimal kernels appears in Donoho and Liu (1991).

A limitation of the renormalization arguments given here is the fact that they refer to the estimation of single linear functionals $T(f)$. For estimating the whole object $E_f\|\hat{f} - f\|_2^2$, rates of convergence arguments, such as those of Birgé (1983), Bretagnolle and Huber (1979), Boyd and Steele (1978), Efroimovich and Pinsker (1982), Ibragimov and Hasminskii (1981), have a different character than the arguments in the case of a single functional. Donoho (1990), Donoho and Johnstone (1991) and Low (1991) give examples where rates in estimating the whole object can also be derived by renormalization ideas. In another direction, Donoho and Kooperberg (1990) describe the possibilities and limitations of renormalization ideas for estimating nonlinear functionals $T(f)$.

2. The optimization problem. We first turn to an apparently special observation scheme. Suppose that K is a linear operator taking functions $f(x)$ into functions $g(t)$, both functions with arguments in R^d and suppose that W is a Brownian sheet, that is, the integral of a standard white noise. We observe a process Y characterized by

$$(1) \qquad Y(dt) = (Kf)(t)\,dt + \varepsilon W(dt), \qquad t \in R^d.$$

Roughly speaking, Y consists of measurements of Kf with added white noise. We suppose that we wish to estimate $T(f)$ and we have knowledge that $f \in \mathscr{F}$, a convex class of functions. The above setup we term the Gaussian experiment $(T, K, \mathscr{F}, W)$.

Below, we will relate this white noise model to a variety of problems in statistics and other fields: density estimation (Sections 4 and 5); nonparametric regression (Section 6); tomography (Section 4); signal recovery (Section 7). In all these cases, the problems reduce asymptotically to a Gaussian experiment of the above form, with K and ε chosen appropriately. The literature on this convergence to Gaussian experiments is vast and we just mention a few articles that make use of it in some way; see Brown and Low (1990), Donoho and Liu (1991), Donoho and Low (1990), Efroimovich and Pinsker (1981, 1982), Low (1992) and Nussbaum (1985).

We are specifically interested in explicit relations between the minimax risk $R(n)$ for estimation of T based on n observations in a model such as density estimation or regression and the minimax risk $R^*(\varepsilon)$ for estimating T from the observations (1). Such relations have been developed in Donoho and Liu (1991) and Donoho and Low (1990), who show that in certain cases we have

$$R(n) \sim R^*(\varepsilon_n), \qquad n \to \infty,$$

provided ε_n is calibrated appropriately, generally as $\varepsilon_n = c/\sqrt{n}$; in other cases their arguments show that at least

$$R^*(c\varepsilon_n) \le R(n) \le R^*(C\varepsilon_n).$$

Such relations immediately reduce rate calculations for $R(n)$ to rate calculations for $R^*(\varepsilon)$. [See also Sections 4 and 6 below.]

Rate calculations for $R^*(\varepsilon)$ are made easy by the considerable body of results on the white noise model, beginning with Ibragimov and Hasminskii

(1984) and Donoho and Liu (1991), both in the case $K = I$, the identity operator. Donoho (1989) studies Gaussian experiments of the above type, and shows that if we define the modulus of continuity of T over $\mathcal{F}$ by

$$\omega(\varepsilon) = \sup\{|T(f_1) - T(f_0)|: \|Kf_1 - Kf_0\|_2 \le \varepsilon,\ f_i \in \mathcal{F}\},$$

where $\|\cdot\|_2 = \|\cdot\|_{L_2(R^d)}$, then

$$\tfrac{1}{4}\omega^2(\varepsilon) \le R^*(\varepsilon) \le \omega^2(\varepsilon),$$

so that the rate in turn reduces to calculation of the modulus of continuity of T over $\mathcal{F}$. Moreover, the exact minimax risk, among affine procedures, of estimating T from data (1) over the class $\mathcal{F}$ is

$$R_A^*(\varepsilon) = \sup_\delta \frac{\omega^2(\delta)\varepsilon^2}{4\varepsilon^2 + \delta^2}. \tag{2}$$

[In the special case where $K = I$, the identity, and $\mathcal{F}$ is centrosymmetric, so that $f \in \mathcal{F}$ implies $-f \in \mathcal{F}$, there is a formula for the linear minimax risk due to Ibragimov and Hasminskii (1984), which, although it does not mention the modulus of continuity, may be seen to be a special case of this formula. Again in the case $K = I$, but without assuming symmetry, the formula given above is established in Donoho and Liu (1991).]

Now if $\mathcal{F}$ is a centrosymmetric class of functions, a further reduction is possible and we get

$$\omega(\varepsilon) = 2\sup T(f) \quad \text{subject to} \quad \|Kf\|_2 \le \varepsilon/2 \quad \text{and} \quad f \in \mathcal{F}.$$

Let us now suppose that membership $f \in \mathcal{F}$ is determined by a functional $J_2(f)$, which, roughly, measures the size of a certain derivative of f:

$$\mathcal{F} = \{f: J_2(f) \le C\}.$$

Then the optimization problem posed above can be written as

$$\omega(\varepsilon) = 2\,\mathrm{val}(\mathcal{P}_{\varepsilon/2,C}),$$

where $\mathrm{val}(\mathcal{P}_{\varepsilon,C})$ is the value of the optimization problem

$$(\mathcal{P}_{\varepsilon,C}): \quad \sup J_0(f) \quad \text{subject to} \quad J_1(f) \le \varepsilon \quad \text{and} \quad J_2(f) \le C,$$

with $J_0(f) = T(f)$ and $J_1(f) = \|Kf\|_2$. Thus, in a certain sense issues of optimal rates of convergence reduce to the properties of certain constrained optimization problems.

3. Renormalization.

While for general J_0, J_1 and J_2, one cannot expect the value of $\mathcal{P}_{\varepsilon,C}$ to be easily available, in the cases of interest to us, a certain homogeneity of these functionals with respect to dilation makes the problem solvable.

DEFINITION 1. Let $a > 0$ and $b > 0$; $\mathcal{U}_{a,b}$ denotes the renormalization operator that takes the function $f = f(t)$ with domain R^d into the function $\mathcal{U}_{a,b} f = af(bt)$.

We note that $\mathscr{U}_{a,b}$ is a bijection of the common function spaces $L_p(R^d)$, $W^{m,p}(R^d)$, $C_0^\infty(R^d)$. In fact, $(\mathscr{U}_{a,b})^{-1} = \mathscr{U}_{a^{-1},b^{-1}}$ so that $(\mathscr{U}_{a,b}\colon a > 0, b > 0)$ is a group of transformations of the measurable functions with domain R^d.

DEFINITION 2. The functional J is homogeneous with dilation exponent s if, for every $f \in \mathrm{Dom}(J)$,

$$J(\mathscr{U}_{a,b} f) = ab^s J(f).$$

Homogeneous functionals occur naturally in analysis; we record a few examples here. The most basic is differentiation at 0. Let $\mathbf{i} = (i_1, \dots, i_d)$ denote a multiindex, let $|\mathbf{i}| = i_1 + \cdots + i_d$ denote its order and set $(D^{\mathbf{i}} f) = (\partial^{|\mathbf{i}|}/(\partial t_1^{i_1} \cdots \partial t_d^{i_d}))f$. Then

$$(D^{\mathbf{i}} \mathscr{U}_{a,b} f)(0) = ab^{|\mathbf{i}|}(D^{\mathbf{i}} f)(0)$$

and so for the functional $J(f) = (D^{\mathbf{i}} f)(0)$ we have exponent $s = |\mathbf{i}|$. For the L_p norm, we have

$$\|\mathscr{U}_{a,b} f\|_p = ab^{-d/p}\|f\|_p \tag{3}$$

and so the exponent is $s = -d/p$.

Homogeneous operators also occur naturally.

DEFINITION 3. A linear operator L which takes measurable functions with domain R^d into measurable functions on R^d is said to be homogeneous, with dilation exponents d and e, if

$$L\mathscr{U}_{a,b} = \mathscr{U}_{ab^d, b^e} L.$$

Of course, differentiation is homogeneous:

$$D^{\mathbf{i}} \mathscr{U}_{a,b} f = \mathscr{U}_{ab^{|\mathbf{i}|}, b}\, D^{\mathbf{i}} f;$$

this implies that every Sobolev functional

$$N_{m,p}(f) = \left(\sum_{|\mathbf{i}| = m} w_{\mathbf{i}} \|D^{\mathbf{i}} f\|_p^p \right)^{1/p}$$

(here the weights $w_{\mathbf{i}} > 0$) is homogeneous, with dilation exponent $s = m - d/p$. One also sees that the Lipschitz-α seminorm

$$L_{m,\alpha}(f) = \max_{|\mathbf{i}|=m} \sup_{u,t} \frac{|D^{\mathbf{i}} f(t) - D^{\mathbf{i}} f(u)|}{|t-u|^\alpha}$$

is homogeneous, with exponent $s = m + \alpha$.

The Fourier transform $\hat{f}(\lambda) = \int e^{-i\langle \lambda, t\rangle} f(t)\, dt$ is also homogeneous:

$$(\mathscr{U}_{a,b} f)^\wedge = \mathscr{U}_{ab^{-d}, b^{-1}} \hat{f};$$

this implies that the functional

$$M_{m,p}(f) = \left(\int_{R^d} |\hat{f}(\lambda)|^p |\lambda|^{mp} \, d\lambda \right)^{1/p}$$

is homogeneous, with exponent $s = m - d + d/p$.

We now turn to more specific cases. The first is the Riesz transform [see Stein (1971) and Ziemer (1990)]:

$$(R_\alpha f)(t) = \frac{1}{\Gamma(\alpha)} \int \frac{f(u)}{|t-u|^{d-\alpha}} \, du.$$

From the Fourier transform of its kernel, we get

$$(R_\alpha f)^\wedge(\lambda) = \hat{f}(\lambda)|\lambda|^{-\alpha}.$$

It follows that the Riesz transform is homogeneous:

$$R_\alpha \mathscr{U}_{a,b} = \mathscr{U}_{ab^{-\alpha},b} R_\alpha \tag{4}$$

and in particular that the functional $\|R_\alpha f\|_2$ has exponent $s = -\alpha - d/2$.

Finally, we consider the Radon transform for functions on R^2; let

$$(P_\theta f)(u) = \int_{-\infty}^{\infty} f(u\cos(\theta) - s\sin(\theta), u\sin(\theta) + s\cos(\theta)) \, ds.$$

Define the functional P by

$$P^2(f) = \int_0^\pi \int_{-\infty}^{\infty} (P_\theta f)^2(u) \, du \, d\theta.$$

The one-dimensional Fourier transform of $(P_\theta f)(\cdot)$ is related to the two-dimensional Fourier transform of f by the so-called Projection-slice theorem (Deans, 1983) which says that

$$(P_\theta f)^\wedge(v) = \hat{f}((v\cos(\theta), v\sin(\theta))). \tag{5}$$

Using this we have

$$P^2(f) = \frac{1}{2\pi} \int_{R^2} \frac{1}{|\lambda|} |\hat{f}(\lambda)|^2 \, d\lambda,$$

and so P is homogeneous with exponent $s = -3/2$.

Table 1 presents a list of functionals, classified as candidates for J_0, J_1 and J_2, and their homogeneity indices.

We now justify our interest in homogeneity. Let $\mathscr{F}_{\varepsilon,C}$ denote the set of functions feasible for program $(\mathscr{P}_{\varepsilon,C})$. Let $(\mathscr{P}_{1,1})$ be the program defined above, with parameters $C = 1$, $\varepsilon = 1$. Define $a(\varepsilon, C)$ and $b(\varepsilon, C)$ as the solutions to the system of equations

$$ab^{s_1} = \varepsilon,$$
$$ab^{s_2} = C.$$

Then, because of the homogeneity of J_1 and J_2, we have

$$\mathscr{U}_{a,b}\mathscr{F}_{1,1} = \mathscr{F}_{\varepsilon,C} \tag{6}$$

TABLE 1
Homogeneous functionals and their exponents

J	Name	Exponents
J_0	$f(0)$	$s_0 = 0$
	$(D^{\mathbf{i}}f)(0)$	$s_0 = \lvert\mathbf{i}\rvert$
J_1	$\lVert f\rVert_{L_2}$	$s_1 = -d/2$
	$\lVert R_\alpha f\rVert_2$	$s_1 = -\alpha - d/2$
	$P(f)$	$s_1 = -3/2$
J_2	$N_{m,p}(f)$	$s_2 = m - d/p$
	$M_{m,p}(f)$	$s_2 = m - d + d/p$
	$L_{m,\alpha}(f)$	$s_2 = m + \alpha$

and

$$\mathscr{U}_{a^{-1},b^{-1}}\mathscr{F}_{\varepsilon,C} = \mathscr{F}_{1,1}. \tag{7}$$

In short, this particular $\mathscr{U}_{a,b}$ is one-to-one and onto between $\mathscr{F}_{\varepsilon,C}$ and $\mathscr{F}_{1,1}$. Now the value of $(\mathscr{P}_{\varepsilon,C})$ may be written:

$$\begin{aligned}\sup\{J_0(f)\colon f\in \mathscr{F}_{\varepsilon,C}\} &= \sup\left\{J_0\left(\mathscr{U}_{a,b}(\mathscr{U}_{a,b})^{-1}f\right)\colon f\in \mathscr{F}_{\varepsilon,C}\right\}\\ &= \sup\left\{J_0(\mathscr{U}_{a,b}g)\colon g = (\mathscr{U}_{a,b})^{-1}f,\ f\in \mathscr{F}_{\varepsilon,C}\right\}\\ &= \sup\{ab^{s_0}J_0(g)\colon g\in \mathscr{F}_{1,1}\}.\end{aligned}$$

Hence

$$\operatorname{val}(\mathscr{P}_{\varepsilon,C}) = ab^{s_0}\operatorname{val}(\mathscr{P}_{1,1}).$$

Taking into account the definition of a and b,

$$a = Cb^{-s_2}, \qquad b = (\varepsilon/C)^{1/(s_1-s_2)},$$

we have proved:

LEMMA 1. *Let J_0, J_1 and J_2 be homogeneous of degrees s_0, s_1 and s_2, respectively. Then*

$$\operatorname{val}(\mathscr{P}_{\varepsilon,C}) = C^{1-r}\varepsilon^r\operatorname{val}(\mathscr{P}_{1,1}), \tag{8}$$

where

$$r = \frac{s_0 - s_2}{s_1 - s_2}. \tag{9}$$

Even if we do not know the constant $\operatorname{val}(\mathscr{P}_{1,1})$, equation (8) gives us the complete qualitative behavior of $\operatorname{val}(\mathscr{P}_{\varepsilon,C})$. Of course, if $\operatorname{val}(\mathscr{P}_{1,1}) = +\infty$, it tells us merely that $\operatorname{val}(\mathscr{P}_{\varepsilon,C}) = \infty$ also. From the monotonicity of $\operatorname{val}(\mathscr{P}_{\varepsilon,C})$ in both ε and C, $r < 0$ is possible only if $\operatorname{val}(\mathscr{P}_{1,1}) = \infty$. Also, as the problem arises from calculating the modulus of continuity of a linear functional, if the feasible class $\mathscr{F}_{\varepsilon,C}$ contains at least two elements, $r \le 1$.

Combining the results of the last two sections and noting (2), we have the following.

THEOREM 1. *Let* $J_0(f) = T(f)$, $J_1(f) = \|Kf\|_2$ *and* $\mathscr{F} = \{f: J_2(f) \le C\}$. *Suppose the* J_i *have dilation exponents* s_i *and that*

$$r = \frac{s_0 - s_2}{s_1 - s_2}$$

takes a value in $[0, 1]$. *Then*

$$\omega(\varepsilon) = \mathrm{val}(\mathscr{P}_{1,1})(2C)^{1-r}\varepsilon^r.$$

The minimax risk, among affine estimates, of estimating T *from observations* (1), *is*

$$R_A^*(\varepsilon) = 2^{2-2r}r^r(1-r)^{1-r}\omega^2(\varepsilon)$$

and the minimax risk among all estimates is at least $4/5$ *of this quantity.*

In particular, the existence of uniformly consistent estimates of T from the observations Y is completely equivalent to the assertion that both $r > 0$ and $\mathrm{val}(\mathscr{P}_{1,1}) < \infty$; and if the minimax risk goes to zero at all, it must go at rate ε^{2r}.

4. Applications. The results above immediately identify the rates of convergence for estimating point functionals in the white noise model. Suppose we observe data according to (1) and we know that $N_{m,p}(f) \le 1$, and we are interested in $T(f) = (D^{\mathbf{i}}f)(0)$, with $k = |\mathbf{i}| < m$. Then $J_0(f) = (D^{\mathbf{i}}f)(0)$, so $s_0 = k$, $J_1(f) = \|f\|_2$, so $s_1 = -d/2$ and $J_2(f) = N_{m,p}(f)$ so $s_2 = m - d/p$. Hence the optimal rate is $r = (m - k - d/p)/(m + d/2 - d/p)$. On the other hand, if we varied the smoothness assumption to $L_{m,\alpha}(f) \le 1$, the rate would be $r = (m + \alpha - k)/(m + \alpha + d/2)$. Finally, with the frequency domain constraint $M_{m,p}(f) \le 1$, we would get $r = (m - d - k + d/p)/(m - d + d/p + d/2)$.

The same pattern of argument can treat much more involved problems with essentially the same ease. Suppose we observe the Radon transform of an unknown function f,

$$Y(d\theta, du) = (P_\theta f)(u)\,du + \varepsilon W(d\theta, du) \tag{10}$$

for $\theta \in [0, \pi]$ and $u \in R$. We wish to estimate $T(f) = f(t_0)$ and we assume that $\mathscr{F}$ is the class of f on R^2 which are absolutely continuous and have absolutely continuous partial derivatives of all degrees through $m - 1$ and for which (say) $N_{m,p}(f) \le C$.

First consider the case where $t_0 = 0$. Then $J_0(f) = f(0)$, so $s_0 = 0$; $J_1(f) = P(f)$, so $s_1 = -3/2$ and, finally, $J_2(f) = N_{m,p}(f)$, so $s_2 = m - 2/p$. We conclude that

$$r = \frac{m - 2/p}{m - 2/p + 3/2}. \tag{11}$$

For the more general case where $t_0 \neq 0$ the functional $T(f) = f(t_0)$ is not homogeneous with respect to dilation, but the modulus of continuity is the same as when $t_0 = 0$. To see why, let τ_0 denote the translation operator $\tau_0 f(\cdot) = f(\cdot + t_0)$. Note that J_1 and J_2 are translation invariant. Still letting $J_0(f) = f(0)$ and letting $\mathscr{F}_{\varepsilon, C}$ denote the feasible f for problem $(\mathscr{P}_{\varepsilon, C})$, we note that the modulus of $T(f) = f(t_0)$ is exactly twice the supremum of $J_0(\tau_0 f)$ over $\mathscr{F}_{\varepsilon, C}$. But

$$\begin{aligned} \sup\{J_0(\tau_0 f) \colon f \in \mathscr{F}_{\varepsilon, C}\} &= \sup\{J_0(\tau_0 f) \colon f \in \tau_0 \mathscr{F}_{\varepsilon, C}\} \\ &= \sup\{J_0(g) \colon g \in \mathscr{F}_{\varepsilon, C}\}, \end{aligned}$$

where we used $\tau_0 \mathscr{F}_{\varepsilon, C} = \mathscr{F}_{\varepsilon, C}$ twice. The final quantity is just the case where $t_0 = 0$, where renormalization holds exactly.

These results have concrete applications to problems in statistical estimation, via the following:

Lemma 2. *Let $T(f) = f(0)$ or $T(f) = (D^i f)(0)$ and let $\mathscr{F}$ be a class of smooth functions defined by one of the conditions $N_{m,p}(f) \le C$ or $L_{m,\alpha}(f) \le C$ or $M_{m,p}(f) \le C$. Let $R_A^*(\varepsilon; T, K, \mathscr{F})$ denote the minimax risk for recovery of T from observations* (1). *Let $\mathscr{G}$ be the (convex) class of all densities $g = Kf$, with K a bounded linear operator of L_2, $f \in \mathscr{F}$ and $\|g\|_\infty \le M$. Suppose we have observations $X_1, \dots, X_n$ i.i.d. g, where $g \in \mathscr{G}$. For the minimax risk $R(n, T, \mathscr{G})$ for estimating $T(f)$ from these observations*

$$R(n, T, \mathscr{G}) \asymp R_A^*\left(\frac{1}{\sqrt{n}}; T, K, \mathscr{F}\right).$$

This is proved in Donoho and Low (1990).

We first apply this to density estimation. Let $\mathscr{D}$ be the class of all densities f satisfying $\|f\|_\infty \le M$ and $f \in \mathscr{F}$, where $\mathscr{F}$ is one of the three types of smoothness classes mentioned earlier. We are interested in estimating $T(f) = (D^i f)(0)$ from data $X_1, \dots, X_n$ i.i.d. f. The above lemma applies with $K = I$, the identity operator and $g = f$. We therefore have that the optimal rates for the Gaussian experiments calculated in the first paragraph of this section are also optimal rates for the density estimation problem.

In this way we may recapture optimal rate results of, for example, Stone (1980), who studied density estimation at a point over the class defined by $L_{m,\alpha}(f) \le C$. We also get previously unpublished results, by considering other classes, such as $N_{m,p}(f) \le 1$.

We turn now to an application to an (idealized) form of positron emission tomography [compare Johnstone and Silverman (1990)]. In PET we (ideally) observe (θ_i, u_i), $i = 1, \dots, n$, i.i.d. from a probability density $g(\theta, u)$ which is the Radon transform of a density $f(t)$. We wish to estimate $T(f) = f(t_0)$. Our a priori class $\mathscr{D}$ is the set of densities supported on all of R^d, satisfying $N_{m,p}(f) \le C$ and $\|g\|_\infty \le M < \infty$. Applying the above lemma with $K = P_\theta$, we see that the optimal rate of convergence of the minimax risk to zero is simply n^{-r}, where r is the rate exponent calculated for the corresponding Gaussian experiment. (The same rate holds if we consider only functions and densities supported in the unit disk.) This result on optimal rates at a point from Radon density data appears to be new.

5. Inhomogeneous K. The homogeneity of the Radon transform is rather special. For most operators K, the functional $J_1(f) = \|Kf\|_2$ will not be exactly homogeneous. This is an apparent obstacle to our approach. Actually, however, many such functionals are asymptotically homogeneous; and if we analyze the problem which arises by replacing the original operator K by its homogeneous approximation K^*, we get the right answer.

We describe the idea in the deconvolution problem. Suppose that our observations operator has the form

$$(Kf)(t) = \int k(t-u) f(u)\, du,$$

where the kernel k has Fourier transform $\hat{k}(\lambda)$. Suppose, in addition, that

$$|\hat{k}(\lambda)| \sim A|\lambda|^{-\alpha}, \qquad |\lambda| \to \infty.$$

Thus, at high frequencies, $|\hat{k}(\lambda)|$ behaves like a power law. Equivalently,

$$\frac{|\hat{k}(\lambda)|}{|\hat{R}_\alpha(\lambda)|} \to A \quad \text{as } |\lambda| \to \infty. \tag{12}$$

The operator $K^* = AR_\alpha$ is therefore asymptotically equivalent to K and it renormalizes exactly. Let us suppose $A = 1$.

If, instead of Y we had observed the process Y^* characterized by

$$Y^*(dt) = (R_\alpha f)(t)\, dt + \varepsilon W(dt), \qquad t \in \mathbb{R}^d$$

we would be in an exactly renormalizing situation. $\|R_\alpha f\|_2$ is homogeneous with exponent $-\alpha - d/2$. Hence if our goal were to estimate $T(f) = f(0)$ and our a priori information were $N_{m,p}(f) \le C$, we would have the J_i all homogeneous, with exponents $s_0 = 0$, $s_1 = -\alpha - d/2$ and $s_2 = m - d/p$. The optimal rate of convergence for this problem would be

$$r^* = \frac{m - d/p}{m - d/p + \alpha + d/2}.$$

It turns out that under sufficient regularity of K, $r^* = r$; hence this rate calculation, made assuming that K^*, rather than K, generated our observations, nevertheless applies to the inhomogeneous experiment using K.

THEOREM 2. *Suppose that k is asymptotically equivalent to the Riesz kernel, so that* (12) *holds. Let $\omega(\varepsilon)$ denote the modulus of continuity of T over*

$\mathscr{F} = \{N_{m,p}(f) \le C\}$ *with respect to the distance* $\|Kf_1 - Kf_0\|_2$ *and let* $\omega^*(\varepsilon)$ *denote the modulus with respect to* $\|R_\alpha f_1 - R_\alpha f_0\|_2$.

Define

$$M(\Lambda) = \sup\left\{\frac{|\hat{k}(\lambda)|}{|\hat{R}_\alpha(\lambda)|} : |\lambda| \le \Lambda\right\}.$$

If $M(\Lambda) < \infty$ *for each positive* Λ, *then*

$$\omega(\varepsilon) \ge \omega^*(\varepsilon)(1 + o(1)) \quad \text{as } \varepsilon \to 0. \tag{13}$$

Define

$$m(\Lambda) = \inf\{|\hat{k}(\lambda)| : |\lambda| \le \Lambda\}/|\hat{R}_\alpha(\Lambda)|.$$

If $m(\Lambda) \to 1$ *as* $\Lambda \to \infty$, *then*

$$\omega^*(\varepsilon) \ge \omega(\varepsilon)(1 + o(1)) \quad \text{as } \varepsilon \to 0. \tag{14}$$

The proof is given in Section 9.

For an elementary example, let $d = 1$, and we observe $X_1, \dots, X_n$ i.i.d. g. Suppose the density g is known to be the convolution of a density f, the object of interest, with a mean-1 exponential waiting time density. We suppose that $f \in \mathscr{F}$, where $\mathscr{F}$ is one of our three types of smoothness class, and that $\|g\|_\infty \le M$ for some positive finite constant M. We are interested in recovering $T(f) = f(t_0)$. We may apply Lemma 2 to this situation, and reduce the problem to calculations in the white noise model. As the Fourier transform of the standard exponential density is $(1 + i\lambda)/(1 + \lambda^2)$, we easily see that the modulus is equivalent to $\hat{R}_1$. Hence for the associated renormalizable Gaussian experiment, we get the exponent $r = s_2/(s_2 - 3/2)$, where s_2 is the exponent of the smoothness functional we have chosen. By Lemma 2, the rate in the density deconvolution problem is therefore n^{-r}.

6. Inhomogeneous domains. In a number of practical cases we are interested in functions f defined not for all $t \in R^d$, but only for t in a certain domain D. If D is a standard half-space $\{t_i \ge 0\}$ or orthant $\bigcap_{j=1}^m \{t_{i_j} \ge 0\}$, then $\mathscr{U}_{a,b}$ is a one-to-one mapping of measurable functions on D onto itself. In that case, renormalization continues to work smoothly. But if D is some bounded region of R^d, $\mathscr{U}_{a,b}$ is not domain-preserving. However, renormalization may continue to apply in an asymptotic sense.

For a concrete example, suppose that $d = 1$ and that $D = [-a, a]$ with $0 < a < \infty$. We observe a process Y characterized by

$$Y(dt) = f(t)\,dt + \varepsilon W(dt), \qquad t \in D,$$

where $W(dt)$ is a white noise. Suppose we know that $\int_D |f^{(m)}(t)|^p\,dt \le C^p$ and we wish to estimate $T(f) = f(0)$. This leads immediately to the optimization problem

$$(\mathscr{P}_{\varepsilon,C,D}): \ \sup f(0) \quad \text{subject to} \quad \int_D f^2 \le \varepsilon^2 \quad \text{and} \quad \int_D |f^{(m)}|^p \le C^p.$$

Let $(\mathscr{P}_{\varepsilon,C})$ denote the value of the corresponding problem with D replaced by $(-\infty,\infty)$ throughout. While for finite $\varepsilon > 0$, we must expect that $\mathrm{val}(\mathscr{P}_{\varepsilon,C,D}) \neq \mathrm{val}(\mathscr{P}_{\varepsilon,C})$, there is still asymptotic agreement.

THEOREM 3. *Let $a \in (0,\infty)$ and $C \in (0,\infty)$.*

$$\frac{\mathrm{val}(\mathscr{P}_{\varepsilon,C,D})}{\mathrm{val}(\mathscr{P}_{\varepsilon,C})} \to 1 \quad \text{as } \varepsilon \to 0.$$

The proof is in the Appendix.

It follows that if ω denotes the modulus of continuity of T in the problem with domain D and if ω^* the modulus in the problem with homogeneous domain $(-\infty,\infty)$, then

$$\omega(\varepsilon) \sim \omega^*(\varepsilon) \quad \text{as } \varepsilon \to 0.$$

This fact has been stated, but no proof given, in Donoho and Liu [(1991), Lemma 11].

For an application, consider the problem of random design nonparametric regression. We wish to recover $T(f) = f(0)$ from observations $y_i = f(u_i) + z_i$, with u_i i.i.d. uniform on $(-1/2, 1/2)$ and z_i i.i.d. $N(0,\sigma^2)$ independently of the (u_i). We suppose that $N_{m,p}(f) \le C$ and that $\|f\|_\infty \le M$.

The problem of determining the optimal rate of convergence reduces to determining that in a Gaussian experiment, by the following result of Donoho and Low (1990).

LEMMA 3. *Let T be linear, $\mathscr{F}$ convex and $\|f\|_\infty \le M < \infty$ for every $f \in \mathscr{F}$. Then the minimax risk $R_A(n)$ for estimating T from n observations in the random design nonparametric regression problem is related to the minimax risk $R_A^*(\varepsilon)$ for estimating T from observations $Y(dt) = f(t)\,dt + \varepsilon W(dt)$, $t \in [-1/2, 1/2]$ by*

$$R_A^*(\sigma/\sqrt{n}) \le R_A(n) \le R_A^*(\tau/\sqrt{n}),$$

with $\tau^2 = \sigma^2 + M^2$. If the optimal rate of convergence in the white noise problem is ε^{2r}, then the optimal rate in the nonparametric regression problem is n^{-r}.

This reduces us to rate calculation in a white noise problem with inhomogeneous domain $[-1/2, 1/2]$. Applying Theorem 3 above reduces us to a calculation in a white noise problem with homogeneous domain. Applying renormalization, we get the rate in the white noise problem and hence the optimal rate for nonparametric regression with random design. For the smoothness class $L_{m,\alpha}(f) \le C$, the rate is $r = (m+\alpha)/(m+\alpha+1/2)$. This recovers results of Stone (1980). Other results are also easily available. For smoothness constraint $N_{m,p}(f) \le C$, we get the rate $r = (m - 1/p)/(m - 1/p + 1/2)$.

7. The renormalization heuristic. From the authors' point of view, the most significant contribution of the renormalization idea is to quickly and easily identify a candidate for the optimal rate of convergence in a wide variety of problems. The steps involved in this process are:

1. Replace the model under study by the asymptotically equivalent white noise model.
2. Replace that white noise model by an asymptotically equivalent white noise model which renormalizes exactly.
3. Calculate the optimal rate for the renormalizable model via Theorem 1.
4. Justify the approximations in steps 1 and 2, and prove that the value of $(\mathscr{P}_{1,1})$ is finite.

In many problems of which we are aware, identifying the correspondences in steps 1 and 2 take only a few minutes. One of us (Donoho) has successfully taught steps [1]–[3] from a draft of the present manuscript in a graduate course at Berkeley. Step 4 takes more time and effort; see the proofs of Theorems 2 and 3 in this paper and also the paper Donoho and Low (1990).

In our opinion, the four-step process described above can be used to clarify and simplify the literature on optimal rates of convergence. Many papers on this subject have the flavor of involving lengthy and arcane calculations and the rate emerges only after the dust settles. The reader is left with no way to derive the rate for himself, or study the effect of variations in the assumptions. In contrast, renormalization allows one to validate, in a few moments, the plausibility of such rate results. It also allows the reader to consider variations on the choice of $\mathscr{F}$ and T and see how these affect the optimal rate.

In this section we present a few examples to give the flavor of renormalization-based reasoning in action.

7.1. *Hall's signal recovery model.* Hall (1990) considers a model for recovery of an image from out-of-focus, noisy data. The model is rather involved to state in its original form; but step 1 of our heuristic reduces it to the following problem. Suppose we observe

$$Y(dt) = (k * f)(t)\,dt + \varepsilon W(dt), \qquad t \in R^d, \tag{15}$$

where W is the integral of a standard white noise on R^d and $k * f$ is the convolution product $\int k(\cdot - u) f(u)\,du$. About the out-of-focus filter k we know that in the frequency domain,

$$\hat{k}(\lambda) = \prod_{i=1}^{d} \frac{1}{1 + c|\lambda_i|^{\nu}}$$

and about the function f we know that

$$M_{m,\infty}(f) \le 1.$$

We are interested in recovering $T(f) = f(0)$.

Now $\|Kf\|_2$ is not exactly renormalizable, but $\hat{k}$ is visibly asymptotic to

$$\hat{k}^*(\lambda) = \frac{1}{c^d} \prod_{i=1}^{d} \frac{1}{|\lambda_i|^{\nu}}$$

as $|\lambda| \to \infty$. Now $\|k^* * f\|_2$ is renormalizable, with exponent $s_1 = -d\nu - d/2$. It follows that for recovering $T(f) = f(0)$ from observations $Y(dt) = (k^* * f)(t)\,dt + \varepsilon W(dt)$, the optimal rate will be

$$r^* = \frac{m - d}{m - d/2 + d\nu}.$$

Although k^* is not Riesz, we may expect that a result parallel to Theorem 2 applies in this case and so for recovering $T(f) = f(0)$ from data (15), the optimal rate $r = r^*$. Translating back to Hall's original model requires the calibration $\varepsilon^2 = (\sigma^2/n)$. Under this calibration, we get the prediction that the minimax mean squared error goes to zero at a rate n^{-r^*}. This agrees with the result established by Hall from lengthy calculations. The heuristic allows us to see easily that this *ought* to be the correct rate; and it allows us to calculate rate heuristics under, for example, variation of the smoothness class easily and naturally.

7.2. *Partial deconvolution*. There has been considerable interest in deconvolution problems and associated optimal rates recently [Rice and Rosenblatt (1983), Ritov (1986), Stefanski and Carroll (1990), Carroll and Hall (1988), Fan (1989)]. In general, deconvolution is quite difficult, in the sense that the rate r is typically close to 0. Renormalization ideas allow one to easily explore related problems which are much better behaved. The first author's (Donoho) attention to the general idea of partial deconvolution was stimulated by a Berkeley seminar of Ritov, who first suggested and obtained results for the problem of partial Gaussian deconvolution.

Suppose we are given convolution data as in (15) and we know that $f \in \mathscr{F}$. For the problem of recovering $f(0)$, these data are less useful than we might desire, for the reason that sharp deltalike features in f are blurred out in the observations $k * f$ and so a loss of resolution occurs.

As an alternative to total deconvolution, consider the idea of partial deconvolution. Let $k_h(t) = k(t/h)/h^d$ be the rescaled kernel—we think here of positive h, $h \ll 1$. Consider recovering $T_h(f) = (k_h * f)(0)$. This is a local average of f in the neighborhood of 0 and it is an attempt to estimate the data we would have recovered with an improved instrument, which had a narrower pulse width.

According to the heuristic, suppose that k is actually the Riesz kernel R_α for some α. Then for $J_0 = T_h$, we get $s_0 = -\alpha$; for $J_1 = \|k * f\|_2$, we have $s_1 = -\alpha - d/2$; and for $\mathscr{F}$ defined by the constraint $N_{m,p}(f) \le 1$, we have $s_2 = m - d/p$. It follows that the rate for partial deconvolution is

$$r^*_{\text{partial}} = \frac{m + \alpha - d/p}{m + \alpha - d/p + d/2}.$$

In contrast the rate for total deconvolution would be

$$r^*_{\text{total}} = \frac{m - d/p}{m + \alpha - d/p + d/2}.$$

If the problem is highly ill-posed, so that α is large, then the rate of convergence for partial deconvolution can be considerably better than the rate for total deconvolution. Thus a reasonable response to the ill-posedness of deconvolution would be to demand compression, rather than removal, of the convolution filter.

7.3. *Nonrenormalization*. Of course, not every problem admits of renormalization or even approximate renormalization. The canonical example is Gaussian deconvolution. Suppose we observe

$$Y(dt) = (\phi * f)(t)\,dt + \varepsilon W(dt), \qquad t \in R^1,$$

where ϕ denotes the Gaussian density $(1/\sqrt{2\pi})e^{-t^2/2}$. Because $\hat{\phi}(\lambda) = e^{-\lambda^2/2}$, ϕ is inequivalent to any Riesz kernel or other renormalizing kernel. It turns out that no algebraic rate ε^{2r} typically holds for this problem; instead, a logarithmic rate $\log^{2\gamma}(1/\varepsilon)$ obtains [Ritov (1986), Carroll and Hall (1988)]. Other examples where renormalization fails can be constructed by modifying this basic example.

8. Optimal kernels. Renormalization may also be used to derive optimal kernels, that is, kernels of linear estimators which are minimax for mean squared error among linear estimates. Several notions of optimality of kernels have been considered—compare, for example, Gasser and Müller (1979), Müller (1984). An explicit statement of the notion of minimax mean-squared error optimality which we consider here is due originally to Sacks and Ylvisaker (1981).

In the general, not necessarily renormalizing case, Donoho (1989) shows that if we put

$$\varepsilon_0 = \arg\max_{\delta} \frac{\omega^2(\delta)}{4\varepsilon^2 + \delta^2},$$

and if $\mathscr{F}$ is centrosymmetric, with f_{ε_0} solving the problem

$$\sup T(f) \quad \text{subject to} \quad \|Kf\| \le \varepsilon_0/2 \text{ and } f \in \mathscr{F},$$

then

$$\psi_\varepsilon(t) = c_0 \frac{2\omega(\varepsilon_0)}{\varepsilon_0^2}(Kf_{\varepsilon_0})(t),$$

is an optimal kernel, where

$$c_0 = \frac{\varepsilon_0^2}{\varepsilon_0^2 + 4\varepsilon^2};$$

this kernel is optimal in the sense that

$$\hat{T}_\varepsilon(Y) = \int \psi_\varepsilon(t) Y(dt)$$

is minimax linear for T at noise level ε.

Using the scaling relations of this paper, if f_1 denotes a solution to $(\mathscr{P}_{1,1})$, then $f_{\varepsilon_0} = \mathscr{U}_{a,b} f_1$ for appropriate a, b; moreover,

$$\varepsilon_0 = 2\sqrt{\frac{r}{1-r}}\,\varepsilon, \qquad c_0 = r, \tag{16}$$

and so

$$\psi_\varepsilon(t) = 2r\,\mathrm{val}(\mathscr{P}_{1,1}) C^{1-r}\left(\sqrt{\frac{r}{1-r}}\,\varepsilon\right)^{r-2} \cdot (K\mathscr{U}_{a,b} f_1)(t). \tag{17}$$

In short, the optimal kernel at any given noise level derives from the solution to the single optimization problem $(\mathscr{P}_{1,1})$ and rescaling.

In the special case where $K = I$, the identity operator, this fact has been noticed before. Donoho and Liu [(1991), Section 4.3] show that the optimal kernels in a family of estimation problems derive from scaling and dilation of the extremal functions for the Kolmogorov–Sobolev–Landau–Hardy inequalities between functions and their derivatives. Those extremal functions derive from a problem of the form $(\mathscr{P}_{1,1})$ with certain homogeneous functionals.

8.1. *Kernel for deconvolution.* We give specifics for $d = 1$ only. Suppose we observe a process Y characterized by

$$Y(dt) = (R_\alpha f)(t)\,dt + \varepsilon W(dt), \qquad t \in R, \tag{18}$$

and that $M_{m,2}(f) \le C$. We wish to recover $T(f) = f(0)$. The corresponding functionals renormalize exactly. The exponents of these functionals have been derived in Section 3 and are recorded in Table 1. The problem $(\mathscr{P}_{1,1})$ becomes

$$\sup \frac{1}{2\pi}\int \hat{f}(\lambda)\,d\lambda \quad \text{subject to} \quad \int |\lambda|^{-2\alpha} |\hat{f}|^2(\lambda)\,d\lambda \le 2\pi$$

$$\text{and} \quad \int |\lambda|^{2m} |\hat{f}|^2(\lambda)\,d\lambda \le 2\pi.$$

The three functionals are invariant under the reflection $f(t) \to f(-t)$; by convexity, a solution f_1 may be taken to be even; hence its Fourier transform $\hat{f}$ will be even and real. A standard variational argument says that a solution must satisfy

$$\int \hat{h}(\lambda)\,d\lambda \le 0$$

whenever

$$\int |\lambda|^{-2\alpha} \hat{f}_1(\lambda) \hat{h}(\lambda)\,d\lambda \le 0 \quad \text{and} \quad \int |\lambda|^{2m} \hat{f}_1(\lambda) \hat{h}(\lambda)\,d\lambda \le 0$$

for every real, even $\hat{h} \in C_K^\infty$. We conclude that for some nonnegative a_1, b_1, we must have

$$\hat{f}_1(\lambda) = \left(a_1|\lambda|^{-2\alpha} + b_1|\lambda|^{2m}\right)^{-1}.$$

By a rescaling argument, the extremal function must attain both constraints

$$\int |\lambda|^{-2\alpha} \left|\hat{f}_1\right|^2(\lambda)\, d\lambda = 2\pi,$$

$$\int |\lambda|^{2m} \left|\hat{f}_1\right|^2(\lambda)\, d\lambda = 2\pi.$$

Using the tabled definite integrals

$$\int_0^\infty \frac{x^p\, dx}{(a+bx)^2} = a^{-1+p} b^{-(p+1)} \frac{p\pi}{\sin(p\pi)} \tag{19}$$

valid for $p \in (0,1)$ and

$$\int_0^\infty \frac{x^q\, dx}{(a+bx)^2} = a^{-1+q} b^{-(q+1)} \Gamma(q+1)\Gamma(1-q) \tag{20}$$

valid for $q \in (-1,0)$, we get the following conditions:

$$a_1^{-1+p} b_1^{-(p+1)} = \frac{(2m+2\alpha)\sin(p\pi)}{p},$$

$$a_1^{-1+q} b_1^{-(q+1)} = \frac{(2m+2\alpha)\pi}{\Gamma(q+1)\Gamma(1-q)},$$

where $p = (2\alpha+1)/(2m+2\alpha)$ and $q = (-2m+1)/(2m+2\alpha)$. The value of the problem is then

$$\begin{aligned}\operatorname{val}(\mathscr{P}_{1,1}) = \frac{1}{2\pi}\int \hat{f}_1(\lambda)\, d\lambda &= \frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\left(a_1\lambda^{-2\alpha} + b_1\lambda^{2m}\right)} \\ &= \frac{1}{2m+2\alpha} b_1^{-(q+1)} a_1^q \frac{1}{\sin(|q|\pi)}.\end{aligned}$$

To complete the analysis, we derive the optimal kernel. Define

$$\phi_{\alpha,m}(t) = \frac{1}{\pi}\int_0^\infty \cos(t\lambda) \frac{\lambda^{-\alpha}\, d\lambda}{a_1\lambda^{-2\alpha} + b_1\lambda^{2m}}.$$

From the homogeneity (4) of the Riesz transform, we have that the term $R_\alpha \mathscr{U}_{a,b} f_1$ demanded by (17) is the same as $\mathscr{U}_{ab^{-\alpha},b}\phi_{\alpha,m}$. With additional calculations we arrive at:

THEOREM 4. *The optimal kernel for recovery of $f(t_0)$ from Riesz data* (18) *for objects known to have m L_2 derivatives is*

$$\psi_\varepsilon(t) = \left(\mathscr{U}_{\gamma,\delta}\phi_{\alpha,m}\right)(t)$$

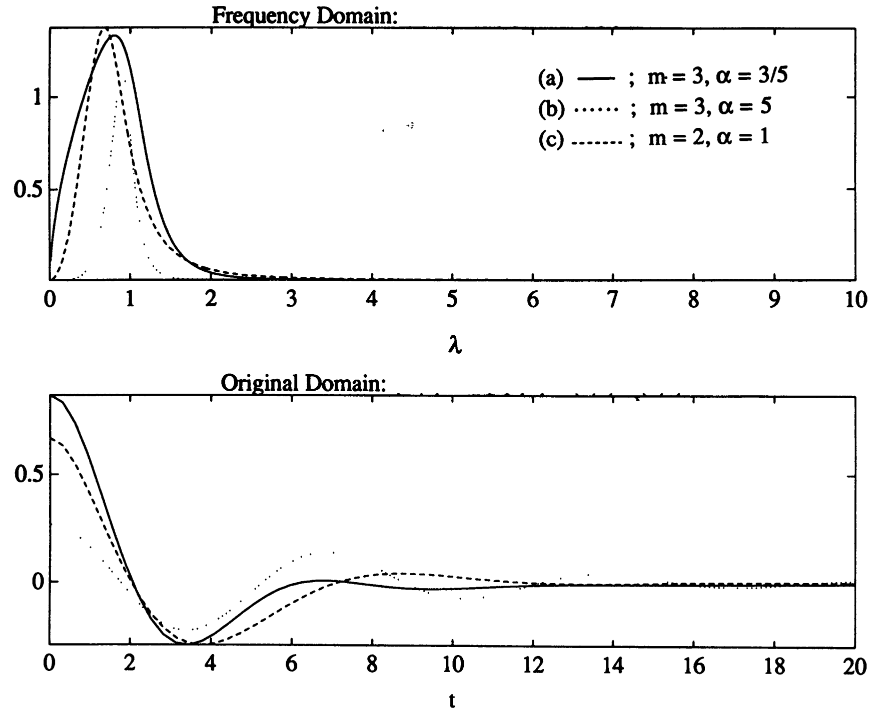

FIG. 1. *Optimal kernels for deconvolution of Riesz transform.*

with scale factors

$$\gamma = 2r \operatorname{val}(\mathscr{P}_{1,1}) \left(\sqrt{\frac{m - 1/2}{\alpha + 1/2}} \frac{\varepsilon}{C} \right)^{-(1+\alpha)/(m+\alpha)}$$

and

$$\delta = \left(\sqrt{\frac{m - 1/2}{\alpha + 1/2}} \frac{\varepsilon}{C} \right)^{-1/(m+\alpha)}.$$

In Figure 1, we display the kernel $\phi_{\alpha, m}$ for three combinations of α and m.

8.2. *Kernel for tomography.* We now assume that f is a function on R^2 and we observe Radon data (10). We wish to estimate $T(f) = f(t_0)$ and assume that f belongs to the class $\mathscr{F}$ of functions on R^2 which are absolutely continuous and have absolutely continuous partial derivatives of all degrees through $m - 1$ and for which $N_{m,2} \leq C$, where the weights $w_i = m!/$

$(i_1!(m - i_1)!)$. The corresponding problem $(\mathscr{P}_{1,1})$ is then

$$\sup(2\pi)^{-2} \int e^{i\langle\lambda, t_0\rangle} \hat{f}(\lambda)\, d\lambda \quad \text{subject to} \quad \int |\lambda|^{-1} |\hat{f}|^2(\lambda)\, d\lambda \le 2\pi,$$

$$\int |\lambda|^{2m} |\hat{f}|^2(\lambda)\, d\lambda \le (2\pi)^2.$$

Letting $\hat{f}_1^* = e^{-i\langle\lambda, t_0\rangle}\hat{f}_1$, we get that f_1^* does not depend on the choice of t_0 and, by a variational argument, must satisfy

$$\hat{f}_1^*(\lambda) = \left(a_1|\lambda|^{-1} + b_1|\lambda|^{2m}\right)^{-1},$$

where a_1 and b_1 are nonnegative constants which satisfy

$$\int |\lambda|^{-1} |\hat{f}_1|^2(\lambda)\, d\lambda = 2\pi,$$

$$\int |\lambda|^{2m} |\hat{f}_1|^2(\lambda)\, d\lambda = (2\pi)^2,$$

or equivalently,

$$\int_0^\infty \left(a_1 r^{-1} + b_1 r^{2m}\right)^{-2} dr = 1,$$

$$\int_0^\infty r^{2m+1}\left(a_1 r^{-1} + b_1 r^{2m}\right)^{-2} dr = 2\pi.$$

The tabled integrals (19)–(20) allow us to reduce this to the system

$$a_1^{-1+p} b_1^{-(p+1)} = \frac{(4m+2)\sin(p\pi)}{p},$$

$$a_1^{-1+q} b_1^{-(q+1)} = \frac{(2m+1)}{\Gamma(q+1)\Gamma(1-q)},$$

where $p = 3/(2m+1)$ and $q = (-2m+2)/(2m+1)$. The value of the problem is then

$$\operatorname{val}(\mathscr{P}_{1,1}) = \frac{1}{(2\pi)^2} \int \hat{f}_1^*(\lambda)\, d\lambda = \frac{1}{2\pi} \int_0^\infty \frac{dr}{a_1 r^{-1} + b_1 r^{2m}}$$
$$= \frac{1}{2m+1} b_1^{-(q+1)} a_1^q \frac{1}{2\sin(|q|\pi)}.$$

To get the optimal kernel, we use the projection-slice theorem (5) to simplify computation of $P_\theta \mathscr{U}_{a,b} f_1$.

LEMMA 4. *Let* $\hat{f}(\lambda) = \rho(|\lambda|)e^{-i\langle\lambda, t_0\rangle}$. *Then*

$$(P_\theta \mathscr{U}_{a,b} f)(u) = (\mathscr{U}_{ab^{-1},b}\varphi)(u - t_{0,1}\cos(\theta) - t_{0,2}\sin(\theta)),$$

where

$$\varphi(u) = \frac{1}{\pi}\int_0^\infty \rho(v)\cos(vu)\, dv.$$

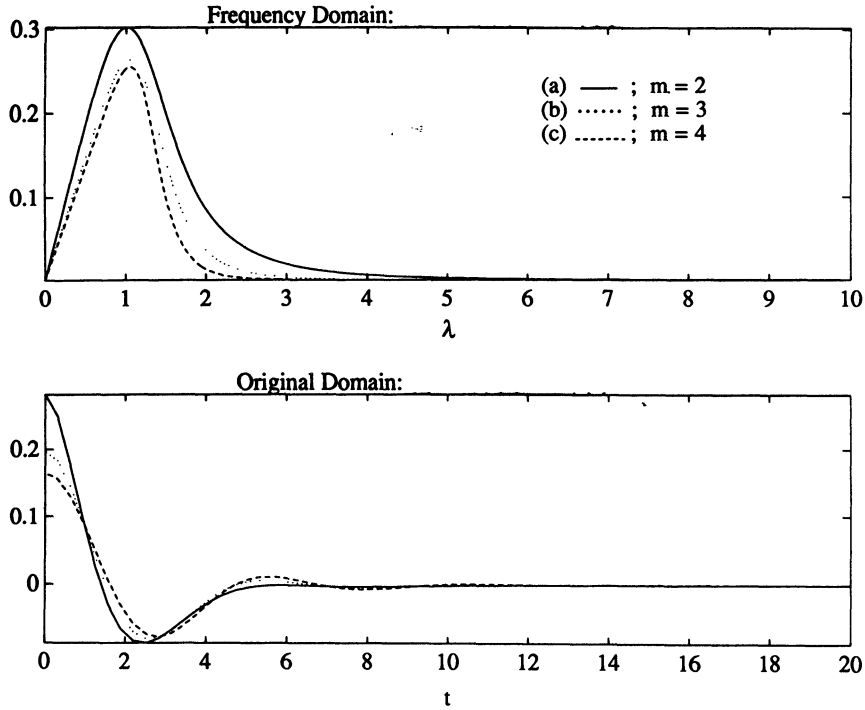

FIG. 2. *Optimal kernel* φ_m *for tomography.*

We omit the proof. Applying this lemma, we define the auxiliary 1-dimensional function

$$\varphi_m(u) = \frac{1}{\pi}\int_0^\infty \frac{\cos(uv)}{a_1 v^{-1} + b_1 v^{2m}}\,dv.$$

THEOREM 5. *The optimal kernel for recovery of* $f(t_0)$ *from Radon data* (10) *is*

$$\psi_\varepsilon(\theta, u) = (\mathscr{U}_{\gamma,\delta}\varphi_m)(u - t_{0,1}\cos(\theta) - t_{0,2}\sin(\theta))$$

for

$$\gamma = 2r\,\mathrm{val}(\mathscr{P}_{1,1})\left(\sqrt{\frac{2}{3}(m-1)}\,\frac{\varepsilon}{C}\right)^{-2/(m+1/2)},$$

$$\delta = \left(\sqrt{\frac{2}{3}(m-1)}\,\frac{\varepsilon}{C}\right)^{-1/(m+1/2)}.$$

The theorem shows that for each fixed θ, the optimal kernel has the form of a rescaled, dilated version of φ_m, which is translated by the value $t_{0,1}\cos(\theta) + t_{0,2}\sin(\theta)$. In tomographic terms, this is an instance of filtered backprojection and we have derived the optimal filter kernel φ_m. See Figures 2 and 3.

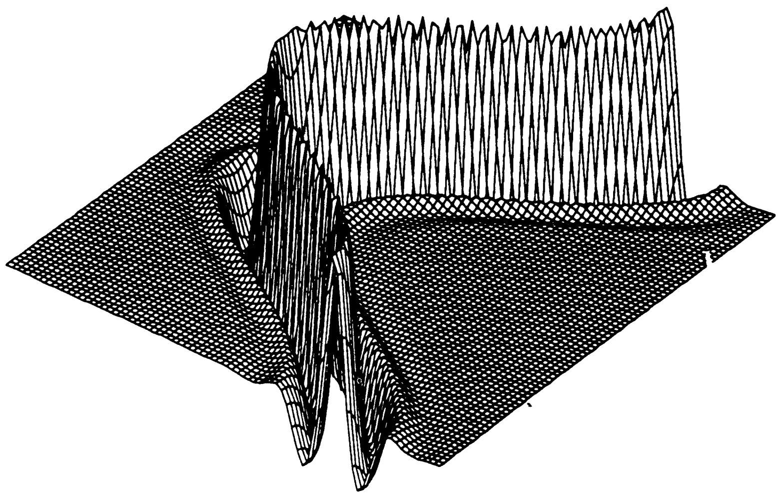

FIG. 3. *Optimal kernel* ψ_ε *for tomography.* $t_0 = (1, 0)$, $\varepsilon = 0.001$, $m = 3$.

8.3. *Boundary kernels.* Heuristic reasoning based on renormalization applies in finding kernels as well. Suppose that D is a convex subset of R^d and we observe

$$Y(dt) = f(t)\,dt + \varepsilon W(dt), \qquad t \in D.$$

We know that $N_{m,p,D}(f) \le 1$, say, and we are interested in recovering $T(f) = f(t_0)$, where $t_0 \in \partial D$. This is a model for recovering a function at the boundary; such boundary problems have attracted considerable interest recently [Gasser and Müller (1979), Rice (1984), Shiau, Wahba and Johnson (1985), Utreras (1986), Eubank and Speckman (1989)].

Without loss of generality, choose coordinates so that $t_0 = 0$. If D is a cone, so that $aD \subset D$ for every $a > 0$, then the problem renormalizes exactly, and one derives rates from Theorem 1, and optimal boundary kernels from solutions to

$$(\mathscr{P}_{1,1,D})\colon \ \sup f(0) \quad \text{subject to} \quad N_{m,p,D}(f) \le 1 \quad \text{and} \quad \int_D f^2 \le 1.$$

If, however, D is not a cone, the problem does not renormalize exactly. Renormalization ideas suggest the following conjecture. Suppose that D^* is the tangent cone to D, that is, the set of all $t \in D$ such that $at \in D$ for all sufficiently small $a > 0$. Then the optimal rate for D ought to be equivalent to the optimal rate for D^* and the optimal kernel for D asymptotically equivalent to the optimal kernel for D^*.

The reader may find it amusing to work out optimal kernels for some particular cones D^*. The computations are often straightforward for $N_{m,2,D}(f)$ with D a half-line, a half-space and an orthant.

We sketch an example for the half-line. We are interested in the problem

$$(\mathscr{P}_{1,1,D}):\quad \sup f(0) \quad \text{subject to} \quad \int_0^\infty (f'')^2 \le 1 \quad \text{and} \quad \int_0^\infty f^2 \le 1.$$

For an appropriate Lagrange multiplier μ, the solution to this problem is also a solution to

$$\inf \int_0^\infty f^2(t) + \mu(f''(t))^2\, dt \quad \text{subject to } f(0) = 1$$

after a rescaling. By a variational argument and integration by parts, f is extremal for the latter problem if for all $h \in C^\infty[0,\infty)$ vanishing at 0 and ∞, we have

$$\int_0^\infty (f(t) + \mu f^{(4)}(t))h(t)\, dt = 0,$$

$$f''h'|_0^\infty = 0.$$

An f satisfying these conditions is the solution of the differential equation $f(t) + \mu f^{(4)}(t) = 0$ on the positive half-line, with boundary conditions $f''(0) = 0$, $f(0) = 1$, $f(\infty) = 0$. From the theory of linear, constant coefficient ordinary differential equations, we are led to the closed form

$$f_\mu(t) = \cos(\omega t)e^{-\omega t}$$

with $\omega = (4\mu)^{-1/4}$. One sees that $\int_0^\infty (f''_\mu(t))^2\, dt = \omega^3/2$ and $\int_0^\infty f_\mu^2 = 3/(8\omega)$; hence picking $\mu = 1/3$ we have that $2^{1/2}(3/4)^{-3/8} f_{1/3}$ is the solution of the original problem $(\mathscr{P}_{1,1,D})$.

8.4. *Nonrenormalizable kernels.* Of course, the lack of renormalization does not prevent one from solving for the minimax kernel. Return to the Gaussian deconvolution example of Section 7.3. Suppose that $N_{m,2}(f) \le 1$; then the solution f_1 to the relevant optimization problem obeys, in the frequency domain,

$$\hat{f}_1(\lambda) = (a\lambda^{2m} + be^{-\lambda^2/2})^{-1}$$

for certain constants a, b; and

$$\hat{\psi}(\lambda) = ce^{-\lambda^2/2}(a\lambda^{2m} + be^{-\lambda^2/2})^{-1}$$

for a certain constant c. Renormalization is a convenience, but not a necessity.

9. Proofs.

9.1. *Proof of Theorem* 2. We divide the proof into two parts.

9.1.1 PROOF OF (14). Let $(\mathscr{P}^*_{\varepsilon,C})$ denote the problem based on the functional $\|R_\alpha f\|_2$ rather than $\|Kf\|_2$. Because $(\mathscr{P}^*_{\varepsilon,C})$ is exactly renormalizable, there exist $a = a(\varepsilon, C)$ and $b = b(\varepsilon, C)$ so that $\mathscr{U}_{a,b}$ maps the solution to $(\mathscr{P}^*_{\varepsilon,C})$ onto the solution of $(\mathscr{P}^*_{1,1})$ and vice versa. Using the same transformation and defining $\hat{K}_\varepsilon(\lambda) = b^\alpha \hat{K}(b\lambda)$, we map a solution of $(\mathscr{P}_{\varepsilon,C})$ into a solution of

$$(\mathscr{P}_{1,1,\varepsilon}): \qquad \sup J_0(f) \quad \text{subject to} \quad \int \left|\hat{K}_\varepsilon(\lambda)\right|^2 \left|\hat{f}(\lambda)\right|^2 d\lambda \le 1 \quad \text{and} \quad N_{m,p}(f) \le 1.$$

As

$$\omega(\varepsilon) = (2C)^{1-r}\varepsilon^r \operatorname{val}(\mathscr{P}_{1,1,\varepsilon})$$

and

$$\omega^*(\varepsilon) = (2C)^{1-r}\varepsilon^r \operatorname{val}(\mathscr{P}^*_{1,1}),$$

the desired relation (14) follows from

$$\limsup_{\varepsilon \to 0} \operatorname{val}(\mathscr{P}_{1,1,\varepsilon}) \le \operatorname{val}(\mathscr{P}^*_{1,1}). \tag{21}$$

To prove (21), note that our hypotheses imply that for each $\delta > 0$, there is $\Lambda(\delta)$ for which

$$m(\lambda) \ge (1-\delta), \qquad |\lambda| > \Lambda,$$

and also

$$\left|\hat{K}(\lambda)\right| \ge |\lambda|^{-\alpha}(1-\delta), \qquad |\lambda| > \Lambda.$$

Define

$$\rho_\delta(\lambda) = \min(|\delta|^{-\alpha}(1-\delta), |\lambda|^{-\alpha}(1-\delta)).$$

Now $b \to \infty$ as $\varepsilon \to 0$, so for $\varepsilon < \varepsilon_0$, we have $b\delta > \Lambda$ and so

$$\left|\hat{K}_\varepsilon(\lambda)\right| \ge \rho_\delta(\lambda) \quad \text{for all } \lambda$$

as soon as $\varepsilon < \varepsilon_0(\delta)$. It follows that for $\varepsilon < \varepsilon_0(\delta)$, then

$$\int \left|\hat{K}_\varepsilon(\lambda)\right|^2 \left|\hat{f}(\lambda)\right|^2 d\lambda \ge \int \rho_\delta^2(\lambda) \left|\hat{f}(\lambda)\right|^2 d\lambda. \tag{22}$$

Now define the optimization problem

$$(\mathscr{Q}_\delta): \qquad \sup J_0(f) \quad \text{subject to} \quad \int \rho_\delta^2(\lambda) \left|\hat{f}(\lambda)\right|^2 d\lambda \le 1 \quad \text{and} \quad N_{m,p}(f) \le 1.$$

As (22) implies that every function feasible for $(\mathscr{P}_{1,1,\varepsilon})$ is also feasible for $(\mathscr{Q}_\delta)$, we have $\operatorname{val}(\mathscr{P}_{1,1,\varepsilon}) \le \operatorname{val}(\mathscr{Q}_\delta)$, $\varepsilon \le \varepsilon_0$.

The lemma to follow shows that $\operatorname{val}(\mathscr{Q}_\delta) \to \operatorname{val}(\mathscr{P}^*_{1,1})$ as $\delta \to 0$. The desired relation (21) follows.

LEMMA 5. $\operatorname{val}(\mathscr{Q}_\delta)$ *is an increasing function of* δ, *hence* $\lim_{\delta \to 0} \operatorname{val}(\mathscr{Q}_\delta)$ *exists; it equals* $\operatorname{val}(\mathscr{P}^*_{1,1})$.

PROOF. Let $J_{1,\delta}(f)^2 = \int \rho_\delta^2(\lambda)|\hat{f}(\lambda)|^2\,d\lambda$. Note that $J_{1,\delta}(f)$ is decreasing in δ for each f. Also, note that given $J_{1,\delta}(f) \le 1$ and $N_{m,p}(f) \le 1$, there exists an absolute constant B such that $\|f\|_2 \le B$. This is most easily seen for $N_{m,2}(f)$.

Let f_δ be a solution to problem $(\mathscr{Q}_\delta)$ and let $\mathscr{F}_\delta$ be the feasible set. Then $\mathscr{F}_\delta$ is a norm-closed, norm-bounded and convex subset of L_2. Let $(f_{1/n})_{n\ge 8}$ be a sequence of solutions of $\mathscr{Q}_{1/n}$. As $\mathscr{F}_{1/8}$ is thus weakly compact, we can extract along a subsequence n_k a weak limit f_0. As $J_{1,\delta}$ is decreasing in δ, $f_0 \in \bigcap_k \mathscr{F}_{1/n_k}$, from which it follows that $J_{1,1/n_k}(f_0) \le 1$ for all k. As $\rho_\delta \to \hat{R}_\alpha$ a.e., Fatou's lemma gives $J_1(f_0) \le \liminf_{k\to\infty} J_{1,1/n_k}(f_0) \le 1$. In other words, f_0 is feasible for $\mathscr{P}^*_{1,1}$.

Notice that on the set $\mathscr{F}_1$, J_0 is a uniformly continuous functional with respect to L_2 distance. Hence, by Lemma 5 of Donoho (1989),

$$\lim_{k\to\infty} T(f_{n_k}) = T(f_0).$$

As $\mathrm{val}(\mathscr{Q}_{1/n_k}) = T(f_{n_k})$, we get

$$\mathrm{val}(\mathscr{Q}_\delta) \to J_0(f)$$

as $\delta \to 0$. But then

$$\mathrm{val}(\mathscr{P}^*_{1,1}) \ge J_0(f_0) = \lim_{\delta\to 0} \mathrm{val}(\mathscr{Q}_\delta)$$

and so

$$\mathrm{val}(\mathscr{P}^*_{1,1}) = \lim_{\delta\to 0} \mathrm{val}(\mathscr{Q}_\delta). \qquad \square$$

9.1.2 PROOF OF (13). It is enough to show that for each $\delta > 0$, then for all $\varepsilon < \varepsilon_0(\delta)$, we have

$$\mathrm{val}(\mathscr{P}^*_{\varepsilon',C}) \le \mathrm{val}(\mathscr{P}_{\varepsilon,C}),$$

where $\varepsilon' = \varepsilon/\sqrt{1+2\delta}$. So, let f^* be a solution to $\mathscr{P}^*_{\varepsilon',C}$. Pick Λ so large that $|\hat{K}(\lambda)|/|\hat{R}_\alpha(\lambda)| \le (1+\delta)^{1/2}$ for $|\lambda| \ge \Lambda$.

$$\begin{aligned}\int |\hat{K}(\lambda)|^2 |\hat{f}^*|^2\,d\lambda &= \int_{-\Lambda}^{\Lambda} |\hat{K}(\lambda)|^2 |\hat{f}^*|^2\,d\lambda + \int_{-\infty}^{-\Lambda} + \int_{\Lambda}^{\infty} |\hat{K}(\lambda)|^2 |\hat{f}^*|^2\,d\lambda \\ &\le M(\Lambda)\int_{-\Lambda}^{\Lambda} |\hat{R}_\alpha(\lambda)|^2 |\hat{f}^*|\,d\lambda \\ &\quad + (1+\delta)\int_{-\infty}^{\Lambda} + \int_{\Lambda}^{\infty} |\hat{R}_\alpha(\lambda)|^2 |\hat{f}^*|^2\,d\lambda.\end{aligned}$$

By renormalization, we may write this as

$$(\varepsilon')^2\Bigg(M(\Lambda)\int_{-\Lambda/b}^{\Lambda/b} |\hat{R}_\alpha(\lambda)|^2 |\hat{f}_1|\,d\lambda + (1+\delta)\int_{-\infty}^{\Lambda/b} + \int_{\Lambda/b}^{\infty} |\hat{R}_\alpha(\lambda)|^2 |\hat{f}_1|^2\,d\lambda\Bigg) \tag{23}$$

with $f_1 = \mathscr{U}_{a,b} f^*$ a solution to $(\mathscr{P}_{1,C})$. Now $\int |\lambda|^{-2\alpha} |\hat{f}^*|^2 \, d\lambda < \infty$, so

$$\int_{-\delta}^{\delta} |\lambda|^{-2\alpha} \left| \hat{f}_1 \right|^2 d\lambda \to 0$$

as $\delta \to 0$. The term in brackets in (23) is smaller than $(1 + \delta + M(\Lambda)o(1))$. For small enough ε,

$$(\varepsilon')^2 (1 + \delta + M(\Lambda)o(1)) \le \varepsilon^2.$$

This implies that f^* is feasible for $\mathscr{P}_{\varepsilon,C}$ and so $\text{val}(\mathscr{P}^*_{\varepsilon',C}) = J_0(f^*) \le \text{val}(\mathscr{P}_{\varepsilon,C})$.

9.2. *Proof of Theorem* 3. Every function feasible for $(\mathscr{P}_{\varepsilon,C})$ is also feasible for $(\mathscr{P}_{\varepsilon,C,D})$; hence

$$\text{val}(\mathscr{P}_{\varepsilon,C,D}) \ge \text{val}(\mathscr{P}_{\varepsilon,C}). \tag{24}$$

We will show in a moment that

$$\text{val}(\mathscr{P}_{\varepsilon,C'}) \ge \text{val}(\mathscr{P}_{\varepsilon,C,D}), \tag{25}$$

where

$$C'(\varepsilon) = C(1 + o(1)) \quad \text{as } \varepsilon \to 0. \tag{26}$$

The result follows from this pair of inequalities and the explicit dependence of $\text{val}(\mathscr{P}_{\varepsilon,C})$ on C stated in Lemma 1.

To establish (25), let ψ be a C^∞ window function satisfying

$$\psi = 1 \quad \text{on } [-a/4, a/4],$$
$$\psi = 0 \quad \text{off } [-a/2, a/2],$$
$$\|\psi\|_\infty = 1.$$

Without loss of generality, suppose that a solution to $(\mathscr{P}_{\varepsilon,C,D})$ exists. Let f_ε be a solution to $(\mathscr{P}_{\varepsilon,C,D})$; and let ψf_ε denote the function equal to $\psi(t) f_\varepsilon(t)$ for $|t| \le a$ and equal to zero elsewhere.

$$T(\psi f_\varepsilon) = T(f_\varepsilon) \quad \text{and} \quad \|\psi f_\varepsilon\|_2 \le \|f_\varepsilon\|_2,$$

where here and throughout the proof, the norms $\|\cdot\|_p$ refer to $L_p[-a, a]$ or to $L_p(-\infty, \infty)$ according to the domain of the argument. It follows that ψf_ε is feasible for $P_{\varepsilon,C'}$, where

$$C' = \left\| (\psi f_\varepsilon)^{(m)} \right\|_p.$$

Hence

$$\text{val}(\mathscr{P}_{\varepsilon,C'}) \ge T(\psi f_\varepsilon) = T(f_\varepsilon) = \text{val}(\mathscr{P}_{\varepsilon,C,D}).$$

Now we bound C':

$$(\psi f_\varepsilon)^{(m)} = \psi f_\varepsilon^{(m)} + c_{m,1} \psi' f_\varepsilon^{(m-1)} + c_{m,2} \psi^{(2)} f_\varepsilon^{(m-2)} + \cdots + \psi^{(m)} f_\varepsilon,$$

with $c_{m,l} = m!/(l!(m-l)!)$; hence

$$\left\| \psi f_\varepsilon^{(m)} \right\|_p \le \|\psi\|_\infty \left\| f_\varepsilon^{(m)} \right\|_p + c_{m,1} \|\psi'\|_\infty \left\| f_\varepsilon^{(m-1)} \right\|_p + \cdots + \left\| \psi^{(m)} \right\|_\infty \|f_\varepsilon\|_p.$$

As $\|\psi\|_\infty = 1$ and $\|f_\varepsilon^{(m)}\|_p \le C$, we get

$$C' \le C + \sum_{l=1}^{m} c_{m,l} \left\| \psi^{(l)} \right\|_\infty \left\| f_\varepsilon^{(m-l)} \right\|_p.$$

Now we apply the lemma following to conclude

$$\| f_\varepsilon^{(m-1)} \|_p \to 0 \quad \text{as } \varepsilon \to 0$$

for $l = 1, \dots, m$, and so

$$C' \le C(1 + o(1)).$$

This completes the proof of (25)–(26).

LEMMA 6. *Suppose that* (f_n) *is a sequence of* $C^\infty[-a, a]$ *functions with* $\| f_n \|_2 \to 0$ *and* $\| f_n^{(m)} \|_p \le C$, *with* $1 \le p \le \infty$; *then* $\| f_n^{(m-l)} \|_p \to 0$ *for* $l = 1, \dots, m$.

The lemma may be obtained by combining compact embedding results in Sobolev space theory with standard inequalities; compare, for example, Adams (1975), Ziemer (1990).

REFERENCES

ADAMS, R. A. (1975). *Sobolev Spaces*. Academic, New York.

BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237.

BOYD, D. W. and STEELE, J. M. (1978). Lower bounds for nonparametric density estimation rates. *Ann. Statist.* **6** 932–934.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137.

BROWN, L. D. and LOW, M. G. (1990). Asymptotic equivalence of nonparametric regression and white noise. Technical report, Cornell Univ.

BROWN, L. D. and LOW, M. G. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression). *Ann. Statist.* **19** 329–337.

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

DEANS, S. R. (1983). *The Radon Transform and Certain of Its Applications*. Wiley, New York.

DONOHO, D. L. (1989). Statistical estimation and optimal recovery. Technical Report 213, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. (1990). Renormalization exponents for minimax filtering. Technical report, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. and JOHNSTONE, I. M. (1990). Wavelets and minimax smoothing of noisy data. Technical report, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. and KOOPERBERG, C. L. (1990). Successes and failures of renormalization in nonlinear settings. Technical report, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.

DONOHO, D. L. and LOW, M. G. (1990). White noise approximation for sampled data. Technical report, Dept. Statistics, Univ. California, Berkeley.

EFROIMOVICH, S. Y. and PINSKER, M. S. (1981). Estimation of square-integrable density on the basis of a sequence of observations. *Problems Inform. Transmission* **17** 50–68.

EFROIMOVICH, S. Y. and PINSKER, M. S. (1982). Estimation of square-integrable probability density of a random variable. *Problems Inform. Transmission* **18** 175–189.

EUBANK, R. and SPECKMAN, P. (1989). A bias reduction theorem with applications in nonparametric regression. Unpublished manuscript.

FAN, J. (1989). Optimal global rates of convergence for nonparametric deconvolution problems. Technical Report 162, Dept. Statistics, Univ. California, Berkeley.

FARRELL, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of density function at a point. *Ann. Math. Statist.* **43** 170–180.

GASSER, TH. and MÜLLER, H. G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques in Curve Estimation. Lecture Notes in Math.* **757** 23–68. Springer, New York.

HALL, P. (1989). On convergence rates in nonparametric problems. *Internat. Statist. Rev.* **57** 45–58.

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

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1984). Nonparametric estimation of the value of a linear functional in a Gaussian white noise. *Theory Probab. Appl.* **29** 1–17.

JOHNSTON, I. M. and SILVERMAN, B. W. (1990). Speeds of estimation in positron emission tomography. *Ann. Statist.* **18** 251–280.

LOW, M. G. (1991). Renormalizing upper and lower bounds for integrated risk in the white noise model. Technical report, Dept. Statistics, Univ. California, Berkeley.

LOW, M. G. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20** 545–554.

MEYER, T. G. (1977). Bounds for estimation of density functions and their derivatives. *Ann. Statist.* **5** 136–142.

MÜLLER, H. G. (1984). Smooth optimum kernel estimators of densities, regression curves, and modes. *Ann. Statist.* **12** 766–774.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.

RICE, J. (1984). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* **13** 893–900.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: Regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

RITOV, Y. (1986). Rates of convergence for Gaussian deconvolution. Unpublished manuscript.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

SACKS, J. and YLVISAKER, N. D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.* **9** 334–346.

SHIAU, J. J., WAHBA, G. and JOHNSON, D. R. (1985). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis. Technical Report 777, Dept. Statistics, Univ. Wisconsin, Madison.

STEFANSKI, L. and CARROLL, R. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184.

STEIN, E. L. (1971). *Singular Integrals and Differentiability Properties of Functions*. Princeton Univ. Press.

STONE, C. J. (1980). Optimum rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

UTRERAS, F. (1986). Boundary effects on convergence rates for Tikhonov regularization. *J. Approx. Theory* **54** 234–249.

WAHBA, G. (1975). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.

ZIEMER, W. (1990). *Weakly Differentiable Functions*. Springer, New York.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302