ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION F FROM A FINITE S-ETC(U)
MAY 81 A F KARR, R J SERFLING
N00014-79-C-0801
ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION $f$
FROM A FINITE SEQUENCE \( \{\text{sgn}(f(t_i^n + x_i^n))\} \).

by

Alan F. Karr and Robert J. Serfling
The Johns Hopkins University

Technical Report No. 341
CNR Technical Report No. 81-4
May 1981

Research supported by the Army, Navy and Air Force under
Reproduction in whole or in part is permitted for any purpose of the United State Government.
ABSTRACT
ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION f
FROM A FINITE SEQUENCE \{sgn(f(t_i) + x_i)\}

Consider reconstructing a function \(f(t)\), \(0 \leq t \leq 1\), from knowledge only of \(\{f(t_i), x_i\}, 1 \leq i \leq n\), where \(x_i = \text{sgn}(f(t_i) + x_i)\), \(1 \leq i \leq n\), and the \(x_i\) are additive "corruptions." Without the components \(x_i\), \(f\) could not be reconstructed. However, for \(f\) continuous and for random uniform noise \(x\), Nasy and Cambanis (1980,1981) show that \(f\) can be consistently estimated almost surely and in mean square as \(n \to \infty\) and establish rates of uniform convergence. Through a somewhat different treatment, in which the approximation of \(f(t_i)\) is identified as a numerical integration problem rather than a statistical problem, we obtain simple exact bounds on the error of estimation, allow the noise \(x\) to be arbitrary (random or deterministic), and deal with the case of \(f\) having discontinuities. The bounds yield substantially improved rates of convergence when the noise values \(x_i\) are from a quasi-random instead of random sequence.

1. Introduction. Consider reconstructing a function \(f(t)\), \(0 \leq t \leq 1\), from knowledge only of \(\{f(t_i), x_i\}, 1 \leq i \leq n\), where \(x_i = \text{sgn}(f(t_i) + x_i)\), \(1 \leq i \leq n\), and the \(x_i\) are additive "corruptions" whose values may or may not be known. In the absence of the \(x_i\), the resulting values \(f_i\) permit only the estimation of \(\text{sgn}(f(t))\), \(0 \leq t \leq 1\), from which, of course, \(f\) cannot be reconstituted. However, with appropriate \(\{x_i\}\), and under minimal assumptions on \(f\), the given "data" permits approximation of \(f\) within any desired accuracy with respect to the \(L^p\)-metrics, \(1 \leq p \leq \infty\), as \(n \to \infty\). In this paper we develop suitable approximating functions \(\tilde{f}(t)\), \(0 \leq t \leq 1\), and establish explicit and useful upper bounds on the approximation error. We also derive convergence rates under relevant conditions.

Such results have application, for example, to communication systems. An unknown continuous-time real signal \(f(t)\) is to be reduced to binary form ("hardlimited"), transmitted, then reconstituted. In this context of nonparametric signal identification, the problem has been previously considered by Nasy and Cambanis (1980, 1981) and Nasy (1981). They assume that \(f\) is continuous and is bounded in magnitude by a known constant \(B\) and show that if the sequence \(f(t_i)\) is deliberately corrupted additively by a sequence of uniform \([-B,B]\) random variables \(x_i\) before hardlimiting, then \(f\) can be estimated consistently almost surely and in mean square as \(n \to \infty\). They also establish rates for these convergences.
The approach of Nayer and Cambanis derives from the fact that, for any \( t_i \), \( f(t_i) \) may be represented as an expectation:

\[ f(t_i) = E_0 \text{sgn}(f(t_i) + x_i), \]

so that an unbiased estimator of \( f(t_i) \) is given by

\[ \hat{f}_0(t_i) = B \text{sgn}(f(t_i) + x_i), \quad 1 \leq i \leq n. \]

They construct \( \hat{f}_0(t) \) elsewhere by interpolation. In their treatment the contribution to the approximation error \( |\hat{f}_0(t) - f(t)| \) from interpolation error is greatly dominated by the contribution due to stochastic variation in the estimates \( \hat{f}_0(t_i) \). It is important, therefore, to seek a reduction of the "stochastic error" component.

In the present paper the approximation of \( f(t_i) \) is identified as a numerical integration problem rather than a statistical problem. In this context, the method of Nayer and Cambanis corresponds to a Monte Carlo approach. Viewing the problem in this fashion, we are able to allow any kind of noise sequence \( \{x_i\} \), random or deterministic, and to obtain general exact upper bounds on the approximation error.

In particular, exploiting modern improvements in the Monte Carlo method, we replace the random sequence \( \{x_i\} \) by a suitable "quasi-random" sequence \( \{x_i\} \). This permits the stochastic error to be replaced by a much smaller "numerical integration error" counterpart, leading to radical improvements in the rates of convergence in the metrics of interest. The approach also handles the case of \( f \) having discontinuities and can be extended to \( f \) defined on a multi-dimensional domain.

A precise formulation of the problem is presented in Section 2 and a suitable approximating function \( \tilde{f} \) is introduced. The approach described above is implemented in Section 3 to obtain bounds on the \( L^p \)-metric approximation error \( \| \tilde{f} - f \|_p \) in terms of the modulus of continuity of the function \( f \) and the discrepancy (from exact uniformity) of the sequence \( \{x_i\} \). For example, for \( f \) Lipschitz of order 1 (Lip 1) and for suitably chosen \( \{x_i\} \), our bounds yield the rate \( O(n^{-1/2}) \) for the convergence of \( \| \tilde{f} - f \|_p \) to 0 as \( n \to \infty \), where \( n \) is the number of evaluations of \( \text{sgn}(f(t_i) + x_i) \). The number \( n \) is an appropriate measure of the "work" involved in calculating the estimator.

In comparison, Nayer and Cambanis (1981) obtain the rate \( O(n^{-1/4+\epsilon}) \), \( \epsilon > 0 \), for \( f \in \text{Lip } 1 \) and \( \{x_i\} \) random, with a nonsequential estimator and, for a sequential estimator of moving average form, Nayar (1981) obtains the rate \( O(n^{-1/3+\epsilon}) \), \( \epsilon > 0 \). Our bounds are applicable to random \( \{x_i\} \) and yield the sharper rate \( O((\log n)^{1/3})^{-1/3} \). Moreover, our estimator is simpler and easier to compute that the moving average estimator of Nayar (1981).

We also deal with the case of \( f \) having discontinuities, for which purpose we consider the \( L^p \)-metrics, \( 1 \leq p < \infty \). Using the above ideas, bounds for \( \| \tilde{f} - f \|_p \) are obtained in Section 4, for \( f \) of bounded variation. For suitable choice of \( \{t_i\} \) and \( \{x_i\} \), the rate \( O((\log n)^{1/3})^{-1/2} \) follows.

There are other ways to use data of the specific form \( \{t_i, x_i\} \) to determine the approximate whereabouts of the function \( f \). Two alternative approaches leading to useful bounds on the \( L^p \)-metric
approximation error are presented in Section 5. However, the numerical integration approach extends to multi-dimensional settings, as discussed in Section 6 along with other comments.

2. Formulation of the problem. Let \( f(t) \), \( 0 \leq t \leq 1 \), be an unknown function satisfying

\[
|f(t)| \leq B, \quad 0 \leq t \leq 1,
\]

with \( B \) known and finite. Define

\[
q(y,u) = B \ln(y + Bu - B), \quad |y| \leq B, \quad 0 \leq u \leq 1.
\]

It is easily checked that \( \int_0^1 q(y,u)du = y, \quad |y| \leq B \). Hence we may represent \( f(t) \) as

\[
f(t) = \int_0^1 q(f(t), u)du.
\]

Our objective is to construct a suitable approximating function \( \tilde{f}(t), 0 \leq t \leq 1 \), on the basis of "data" of the form \([t_i, q_i]\), \( 1 \leq i \leq n \), where \( q_i = q(f(t_i), u_i) \), with \( \{u_i\} \) a sequence in \( [0,1] \), and to give suitable bounds on the error of approximation. A straightforward approach is to estimate \( f(t) \) directly at \( J \) selected points \( t = t_1, \ldots, t_J \) and obtain \( \tilde{f}(t) \) elsewhere by interpolation or by a step function. By (2.2), the estimation of \( f \) at a selected point \( t \) may be viewed as a problem of evaluating an integral, for which we would want to have evaluation of the integrand \( q(f(t), u) \) at \( K \) suitable points \( u = u_1, \ldots, u_K \). In this case the desired data form an array \([t_{jk}, q_{jk}], 1 \leq j \leq J, 1 \leq k \leq K \), with \( q_{jk} = q(f(t_j), u_{jk}) \). However, a further complication must be taken into account: in practice, since the data typically is collected in "real time", \( q(f(t), u) \) might not be observable for more than one value of \( u \) per value of \( t \). To allow for this possibility, we shall assume that the data is an array of the form

\[
\{t_{jk}, q_{jk}, 1 \leq j \leq J, 1 \leq k \leq K\},
\]

where

\[
0 \leq t_{11} \leq \cdots \leq t_{1K} \leq t_{21} \leq \cdots \leq t_{2K} \leq \cdots \leq t_{JK} \leq 1,
\]

and where \( q_{jk} = q(f(t_{jk}), u_{jk}) \). When the inequalities in (2.3b) are strict, the estimation of \( f \) at any specified point is more difficult.

In addition, the double array format (2.3) presents us with a design problem, of choosing the best trade-off between \( J \) and \( K \) for a fixed choice of the number \( n = JK \) of evaluations of \( q(f(t), u) \). As will be seen, this can be viewed as a trade-off between a numerical integration error and an interpolation error.

We confine attention to step function estimators. This does not increase the order of magnitude of the interpolation error in comparison with linear interpolation, for example, and has the
advantages of simplicity and computational ease. Specifically, we assume that \([0,1]\) is divided into intervals \(I_j = [t_{j-1}, t_j), j = 1, \ldots, J\), where \(0 = t_0 < t_1 < \cdots < t_J = 1\) and that the estimator
\[
\hat{f}(t), \quad 0 \leq t \leq 1,
\]
(2.4)

\(\hat{f}(t)\) is constant on each interval \(I_j\).

Moreover, the above \(t_j\) and the \(t_{jk}\) of (2.3) are selected so that

\(t_{jk} \in I_j\) for each \(j\) and \(k\), and the value of \(\hat{f}\) on \(I_j\) is a function only of the relevant portion of the data (2.3), namely \((t_{jk}, u_{jk})\), \(1 \leq k \leq K\). The determination of \(\hat{f}(t)\) within \(I_j\) will now be described.

One natural approach is based on the idea of estimating the integral in (2.2) by a suitable average of values of \(q\). For the data (2.3) this idea leads to the estimator
\[
\hat{f}(t) = \frac{1}{K} \sum_{k=1}^{K} q(t_{jk}, u_{jk}), \quad t \in I_j, 1 \leq j \leq J.
\] (2.5)

We consider this estimator in Sections 3 and 4 and establish bounds on \(|\hat{f} - f|_\infty\) and \(|\hat{f} - f|_p\) (\(1 \leq p < \infty\)), respectively, as discussed in the Introduction.

The data (2.3) may be used in other ways to develop an approximation to a continuous \(f\). Alternatives to the estimator (2.5) are treated in Section 5.

The estimator (2.5) is very efficient computationally. Over each interval \(I_j\) it is necessary to maintain only a running total

\[
\sum_{k=1}^{K} q(t_{jk}, u_{jk}), \quad \text{which becomes } \hat{f}(t), t \in I_j, \text{ when } t \text{ reaches } 1
\]

On the other hand, the moving average estimator of Mason (1981), whose convergence properties are compared in Section 3 to those of the estimator (2.5), is less efficient in that it entails more computations and larger storage requirements. In our notation the estimator of Mason (1981) is a moving average of the form
\[
\hat{f}(t) = \frac{1}{P} \sum_{k=1}^{P} q(t_{jk}, u_{jk})
\]
(2.6)

at selected points \(t_1, \ldots, t_P\) in \([0,1]\), where the \(u_i\) are independent uniform \([0,1]\) random variables, with \(\tilde{f}\) defined elsewhere by linear interpolation. To obtain the estimator at the \(t_i\) requires calculation and storage of \(P\) averages of the form (2.6).

For the estimator (2.5), \(K\), is the number of function evaluations, and is comparable to \(P\) above, but the number of averages calculated and stored is \(J\).

3. Approximation in \(L^p\). Consider the problem as formulated in Section 2. For approximation of a function \(f\) satisfying (2.1) and (2.2) by the function \(\tilde{f}\) given by (2.4) and (2.5), we will derive an upper bound on \(|\tilde{f} - f|_p = \sup_{t} |\tilde{f}(t) - f(t)|\). The bound will involve the modulus of continuity of \(f\), \(\omega(f; \delta)\), \(\sup |f(a) - f(t)| \leq \delta |f(a) - f(t)|\), and the discrepancy of the values \(\{u_{jk}\}\).

For a sequence \(\omega = u_1, u_2, \ldots\) in \([0,1]\), the discrepancy of the initial
segment of length $K$ is defined as

$$D^K_n(\xi) = \sup_{0 < t < 1} \left| \frac{A(t, \xi_n; K; u_1)}{K} - \frac{1}{K} \sum_{k=1}^{K} q(u_k) \right|,$$

where $A(\xi_n; K; u)$ denotes the number of $u_1, \ldots, u_K$ which belong to the set $E$. This quantity measures the departure of $u_1, \ldots, u_K$ from a "uniform" sequence (see Kuipers and Niederreiter (1974) and Niederreiter (1978) for excellent expository discussion of discrepancy and related topics). We now can state the main result of this section.

**Theorem 3.1.** Assume that $|t(t)| \leq B$, $0 \leq t \leq 1$, where $B$ is known, and that the data (2.3) is available. Let $f(t)$, $0 \leq t \leq 1$, be given by (2.4) and (2.5). Then

$$\left\| \int f(t) - f(t) \right\|_{\infty} \leq 2B \max_{1 \leq j \leq 3} \max_{1 \leq t \leq 3} D^K_n(\xi),$$

where $D^K_n(\xi)$ is the discrepancy of the sequence $\xi_1, \ldots, \xi_j$.

The proof of the theorem will make use of an elementary inequality due to Koksma (1942) (see Niederreiter (1978)), which arises in connection with the estimation of an integral of the form $\frac{1}{K} \sum_{k=1}^{K} q(u_k)$ by averages of the form $(1/K) \int \sum_{k=1}^{K} q(u_k)$, where $0 \leq u_1, \ldots, u_K \leq 1$. The approximation error is bounded as follows.

**Lemma 3.1.** If $q$ has finite variation $V(q)$ on $[0,1]$ and $\epsilon = u_1, u_2, \ldots$ is any sequence in $[0,1]$, then for each $K$

$$\left| \int \frac{1}{K} \sum_{k=1}^{K} q(u_k) \right| \leq V(q) D^K_n(\epsilon).$$

**Proof of Theorem 3.1.** Although the integral in (2.2) is of the form $\int q(u) du$, we cannot directly apply (3.2) because the average in (2.5) is not of the form $(1/K) \int q(u_k)$. However, since $q(y,u)$ is nondecreasing in $y$ for each fixed $u$, we have

$$\frac{1}{K} \sum_{k=1}^{K} q(u_k', y, u_k') \leq \int q(y, u) du \leq \frac{1}{K} \sum_{k=1}^{K} q(u_k', y, u_k'),$$

where $m_j$ and $M_j$ are the infimum and supremum, respectively, of $f$ over the interval $\xi_j$. Applying (3.2), noting that $V(q(y, u)) = 2B$ for all $y$ and recalling that $\int q(y, u) du = y$, we obtain

$$\left| \int f(t) - f(t) \right| \leq 2B D^K_n(\xi),$$

and

$$\left| \int \frac{1}{K} \sum_{k=1}^{K} q(u_k', y, u_k') - m_j \right| \leq 2B D^K_n(\xi).$$

By a simple argument using (3.3) - (3.5) and the relation $m_j \leq f(t) \leq M_j$, $t \in \xi_j$, it follows that for $t \in \xi_j$

$$\left| \int f(t) - f(t) \right| \leq 2B D^K_n(\xi),$$

which yields (3.1).
\[ 1 \leq j \leq J. \text{ With these substitutions, the inequality (3.1) becomes} \]
\[ (3.6) \quad \| f - g \|_n \leq C \left( n^{-1} + j^{-r} \right). \]
Suppose, further, that \( f \in \text{Lip} \gamma \) on \([0,1]\). Then (3.6) may be replaced by
\[ (3.7) \quad \| f - g \|_n \leq C \left( n^{-1} + j^{-r} \right), \]
where \( C \) is a constant. Finally, suppose that we choose \( J \) and \( K \) to maximize the rate of convergence of this bound to \( 0 \) as \( n \to \infty \).
We obtain

**Corollary 3.1.** Assume that \( |f| \leq B \) and \( f \in \text{Lip} \gamma \), \( 0 < \gamma \leq 1 \),
on \([0,1]\), and that \( f \) is given by (3.5) with \( J = J_n = n^{1/(r+1)} \),
\( j = j_n, K = K_n = n^{r/(r+1)} \), and \( u_{j,n} = (2\pi)^{-1/2} \). Then
\[ (3.8) \quad \| f - g \|_n = O(n^{-r/(r+1)}), \quad n \to \infty. \]

It should be noted that (3.3) and (3.6) are applicable only to discontinuous \( f \), whereas the results of Maisry and Cambanis (1981)
and Maisry (1981) apply only to continuous \( f \). Of course the bound in (3.6) does not go to zero as \( J, K \to \infty \) if \( f \) is discontinuous
but nevertheless does assure that \( f \) can be reconstructed to within an error \( \| f - g \|_n \to 0 \), \( n \to \infty \).

Returning to the problem of choosing the \( u_{j,n} \); in some cases it is desirable to let the \( u_{j,n} ; 1 \leq j \leq K \), be generated by a single
sequence \( \zeta \) rather than recompute the values for each change of \( K \).

It is known that the fastest rate possible for convergence of
\( u_{j,\zeta} \) to \( 0 \), as \( K \to \infty \), is \( \zeta \), an infinite sequence is \( 0(\log K)^{1/2} \).
This rate is attained by the van der Corput sequence \( \nu \) defined as follows:
\[ v_j = \zeta_{a_j}, \quad j = 0, 1, \ldots, \]
where \( a_0, \ldots, a_j \) are determined by
\[ k = \zeta_{a_j}. \quad j = 0, 1, \ldots, \]
In particular, the discrepancy of the van der Corput sequence satisfies
\[ u_{j,\zeta} \leq \log (K+1), \quad K \to \infty. \]

See Kulper and Wiedner (1974) for details concerning the van
der Corput sequence and other low discrepancy sequences. With use
of the van der Corput sequence, a minor relaxation of the convergence rate in (3.8) results.

However, the rate in (3.8) breaks down radically if the \( u_{j,n} \)
are replaced by independent uniform \([0,1]\) random variables \( u_{j,n} \).
This is because for such a sequence \( \zeta \), the quantity \( G_{j,n}(\zeta) \) is
precisely the Kolmogorov-Smirnov test statistic, which converges
almost surely to \( 0 \) at the exact rate \( O(\log \log K)^{1/2} \) as \( K \to \infty \).
(Off-Chung (1949)). This illustrates the limitations of the Monte
Carlo method in comparison with the so-called quasi-Monte Carlo method based on quasi-random sequences such as the van der Corput sequence (see Niederreiter (1976) for discussion). To derive the counterpart to (3.8) when the \( U_{j_k} \) are random uniform variables instead of equally spaced values, we shall apply an inequality of Dvoretzky, Kiefer and Wolfowitz (1956), which may be stated as follows.

**Lemma 3.2.** Let \( \{x\} \) be a sequence of independent uniform \([0,1]\) random variables. Then for each \( x \)
\[
P\left[ \left| Y_N(x) - x \right| > d \right] \leq Ce^{-2Kd^2},
\]
d\geq 0,
where \( C \) is a universal constant not depending on \( x \).

We then obtain the following.

**Proposition 3.1.** Assume that \( |f| \leq B \) and \( f \in \text{Lip}_Y \) on \([0,1]\), and let
\[
f(t) = \frac{1}{K} \sum_{k=1}^{K} q(t+\lambda_k), \quad t \in I_j,
\]
where \( J = J_n = n^{1/(2y+1)} \) \((\log n)^{-1/2y+1/2} \), the \( I_j \) all have length \( 1/2n \), \( K = K_n = n^{2y/(2y+1)} \) \((\log n)^{1/(2y+1)} \), and where the \( U_{j_k} \) are independent, uniform \([0,1]\) random variables. Then almost surely
\[
\left\| I - f \right\|_2 = O\left( n^{-1/(2y+1)} n^{-y/(2y+1)} \right), \quad n \to \infty.
\]

**Proof.** It follows using (3.9) that the (now random) discrepancy term in (3.1) satisfies
\[
P \left( \max_{1 \leq j \leq n} D_k^n \left( U_{j_k} \right) > d \right) \leq C e^{-2Kd^2}, \quad d > 0,
\]
where \( D_k^n \left( U_{j_k} \right) \) is the discrepancy of the sequence \( U_{j_1}, \ldots, U_{j_K} \).

Let us for the moment take \( J_n = n^{1/(\log n)^{1/2}}, K_n = n^{1/(\log n)^{1/2}}, d_n = A(\log n)^{(1-\delta)/2} - (1-n) \), where \( 0 < \delta < 1 \) and \( 2A^2 \leq n > 1 \).

Then (3.12) yields
\[
\sum_{n=2}^{\infty} P \left( \max_{1 \leq j \leq n} D_k^n \left( U_{j_k} \right) > d_n \right) = O \left( \frac{n^{1/2}}{n^{1/2} - 2A^2} \right).
\]
Therefore, by the Borel-Cantelli lemma,
\[
\max_{1 \leq j \leq n} D_k^n \left( U_{j_k} \right) = O \left( n^{-1/2} \right), \quad n \to \infty,
\]
almost surely. Application of (3.1) and (3.13) with the specified values for \( J_n, K_n \) and \( d_n \) yields (3.11).

For \( y = 1 \) the rate in (3.11) is \( O(\log n)^{1/3} \), which slightly sharpens the rate \( O(\log n) \), \( n \to \infty \), obtained by Masry (1981). However, the corresponding rate \( O(n^{-1/2}) \) in (3.8) represents a
dramatic improvement. Further improvement results if the $v_{jk}$ are chosen adaptively; see Section 5.

4. Estimation in $L^p$ ($1 \leq p \leq \infty$). In this section we investigate the properties of the estimator $\hat{f}$ given by (3.5) with respect to the $L^p$-norms

$$\|f - g\|_p = \left( \int |f(t) - g(t)|^p dt \right)^{1/p}$$

for $1 \leq p < \infty$. The importance of doing so is that we are able to provide rates for the convergence $\|\hat{f} - f\|_p \to 0$ for a wide class of discontinuous functions $f$. This version of the problem, despite its clear significance, seems to have received virtually no previous attention. However, the numerical integration formulation of the problem is extremely natural in this case, so that our methods provide a unified approach to the discontinuous and continuous cases, as well as to the $L^\infty$ and $L^p$ ($1 \leq p \leq \infty$) cases.

We again deal with the estimator $\hat{f}$ given by (2.5); however, in the present exposition it is convenient to represent the points $t_{jk} = t_{jk}^* + v_{jk}(t_j - s_{j-1})$ in the form

$$t_{jk} = t_{jk}^* + v_{jk}(t_j - s_{j-1}),$$

where $0 \leq v_{jk} \leq 1$. Then

$$\hat{f}(t_j) = \frac{1}{K} \sum_{k=1}^K q(f(t_j^* + v_{jk}(t_j - s_{j-1})), u_{jk}), \quad t \in I.$$  \hspace{1cm} (4.1)

The data (2.3) are now represented as $(v_{jk}, u_{jk})$, $1 \leq j \leq 3$, $1 \leq k \leq K$, where $u_{jk} = q(f(t_j^* + v_{jk}(t_j - s_{j-1})), u_{jk}).$

The main result of this section provides an estimate for

$$\|\hat{f} - f\|_p, \; 1 \leq p \leq \infty,$$

in terms of the following notion of two-dimensional discrepancy.

For a two-dimensional sequence $(v, u)$ in $[0,1]^2$, the discrepancy $D^*_p(v, u)$ is given by

$$D^*_p(v, u) = \sup_{I \subset \mathbb{R}} \left| \frac{1}{|I|} \sum_{x \in I} f\left(\frac{x}{|I|} \cdot v, \frac{x}{|I|} \cdot u\right) - \lambda(I)^p \right|,$$

where the supremum is over all intervals $I \subset [0,1] \times [0,1]$ and $\lambda$ denotes Lebesgue measure; see Kuipers and Niederreiter (1974) or Niederreiter (1978) for further details.

**Theorem 4.1** Assume that $|f| \leq R$ and that $f$ is of bounded variation on $[0,1]$, and let $\bar{f}$ be given by (4.1). Then for each $p \in (1, \infty)$,

$$\|\hat{f} - f\|_p \leq C_1 \max_{1 \leq j \leq 3} D^*_p(v_j, u_j)^{1/2} + C_2 \max_{1 \leq j \leq 3} (t_j - s_{j-1})^{1/2},$$

where $D^*_p(v_j, u_j)$ is the two-dimensional discrepancy of the sequence $(v_{j1}, v_{j2}, v_{j3}, u_{j1}, u_{j2}, u_{j3})$. $C_1$ is a constant depending on $f$ only through its variation $V(f)$ and the bound $R$, and $C_2$ is a constant depending on $p$, $V(f)$ and $R$. (These constants are cal-
cated explicitly in the proof.)

PROOF. Define

\[ (4.3) \quad \hat{f}(t) = S_j^{-1} \int_{I_j} f(v) \, dv, \quad t \in I_j, \]

for \( j = 1, \ldots, J \), where \( \delta_j = t_j - t_{j-1} = \lambda(I_j) \). In probabilistic
terminology, the function \( \hat{f} \) is the conditional expectation of \( f \)
in the probability space \( ([0,1], B([0,1], \lambda)) \) given the \( \sigma \)-algebra
generated by the intervals \( I_j \). This implies, for example, that
\( \hat{f} \) is the best approximation to \( f \), in the \( L^2 \)-norm, among functions
constant over each \( I_j \). Let \( \hat{f}_j \) and \( f_j \) be the values of \( \hat{f} \) and \( f \),
respectively, on the interval \( I_j \). We then have

\[ (4.4) \quad \| \hat{f} - f \|_p \leq \| \hat{f} - \hat{f}_j \|_p + \| \hat{f}_j - f \|_p \]
\[ \leq \| \hat{f} - \hat{f}_j \|_\infty + \| \hat{f}_j - f \|_p \]
\[ = \max_{1 \leq |y_j|} | \hat{f}_j - \hat{f}_j | + \| \hat{f}_j - f \|_p \]

In this inequality the first term can be regarded as a numerical
integration error and the second as an interpolation error (or a
smoothing error.)

Now fix \( j \). The term "numerical integration error" is justi-
fied by the observation that

\[ \hat{f}_j = \int_{I_j} f(t_{j-1} + \omega_0), u) \, du. \]

In particular, we then have

\[ (4.5) \quad | \hat{f}_j - f_j | = \left| \int_{I_j} \int_{[0,1]} q(f(t_{j-1} + \omega_k), u) \, dv \, du \right| \]
\[ = 2 \left| \sum_{k=1}^K 1_{E_k} (v_k, u_k) - \int_{I_j} 1_{E_j} (v, u) \, dv \, du \right| \]

where \( 1_E \) denotes the indicator function of the set \( E \) and

\[ E_j = \{(v, u); \ u \in (28)^{-1} \mathbb{R}^3 \hat{f}(t_{j-1} + \omega_0)\}. \]

The multidimensional version of the Koksma inequality (3.2)
- see Kuipers and Niederreiter (1974) - is not applicable because
the function \( (v, u) \rightarrow q(f(t_{j-1} + \omega_0), u) \) is not of bounded varia-
tion in the sense of Hardy and Krause. However, since \( f \) is of
bounded variation on \([0,1]\), the set \( E \) belongs to the class
\( \mathcal{H}_b \) in the context and notation of Niederreiter (1978, pp. 968 -
969), for

\[ b(t) = (1 + v_j) t. \]
where $V_j$ is the variation of the function $v^+(2B)^{-1} (B - f(u_1 + \frac{4}{5} \mathcal{V}_j))$,
which satisfies

\begin{equation}
V_j \leq (2B)^{-1} V(f).
\end{equation}

By Theorems 2.11 and 3.1 and expression (3.1) of Niederreiter (1978),

\begin{equation}
\frac{1}{K} \sum_{k=1}^{K} \int_{[0,1]^2} v_j(x, y) \, dy \, dx \leq D_k(\mathcal{N}_x^{(1)}) \leq 2 \int_{[0,1]^2} \int_{[0,1]^2} (v_j(x, y))^2 \, dx \, dy + \frac{1}{5} (D_k(\mathcal{N}_x^{(1)}))^2,
\end{equation}

where $c(x) = b(x) + 4c = (5 + V_j) x$, $D_k(\mathcal{N}_x^{(1)})$ is the discrepancy

associated with the class $\mathcal{N}_x^{(1)}$ and $v_j = 5 + V_j$. Combining (4.5)-(4.7),

we obtain

\begin{equation}
|e_j - f| \leq 2 \int_{[0,1]^2} (v_j(x, y))^2 \, dx \, dy + \frac{1}{5} (D_k(\mathcal{N}_x^{(1)}))^2 + 11 (D_k(\mathcal{N}_x^{(1)}))^2 \int_{[0,1]^2} (v_j(x, y))^2 \, dx \, dy,
\end{equation}

which establishes the first term of the bound (4.2) with $C_1$ the constant on the right-hand side of (4.8).

Finally, we consider the interpolation (smoothing) error

\begin{equation}
\|e - f\|_P. \quad \text{By expression (11) of Berens and DeVore (1976),}
\end{equation}

\begin{equation}
\|e - f\|_P \leq C (\delta^2 + \delta_2, p(f, 6)),
\end{equation}

where $C$ is a constant depending only on $p$, $\delta_2, p$ is the second-order $L^p$ modulus of smoothness of $f$ (see Berens and DeVore (1976)
or Timan (1963)), and $\delta = \|e_1 - e_2\|_p$ for $e_1(x) = x$. By straightforward calculations (see Example 1 of Berens and DeVore (1976))
one verifies that

\begin{equation}
\delta^2 \leq \frac{1}{2} (p+1)^{-1/p} \max_{1 \leq j \leq 2} \max_{k \neq 0} |t_k - t_{k-1}|,
\end{equation}

while the results from Timan (1963), p. 127, imply that

\begin{equation}
\omega_{2, p}(f; x) \leq (\delta^2)^{1/p},
\end{equation}

where $V(f)$ is the variation of $f$ over $[0,1]$. Consequently, by (4.9)-(4.11),

\begin{equation}
\|e - f\|_P \leq C (2\delta^2 + 2\delta_2, p(f, 6)) + \frac{1}{2} (p+1)^{-1/p} \max_{1 \leq j \leq 2} \max_{k \neq 0} |t_k - t_{k-1}|^{1/2p} + 2^{1/2} V(f)(p+1)^{-1/2p} \max_{1 \leq j \leq 2} \max_{k \neq 0} |t_k - t_{k-1}|^{1/2p},
\end{equation}

\begin{equation}
\leq C_2 \max_{1 \leq j \leq 2} \max_{k \neq 0} |t_k - t_{k-1}|^{1/2p},
\end{equation}

where $C_2 = 2C \max \left\{ \frac{B}{2} (p+1)^{-1/p}, 2^{1/2} V(f)(p+1)^{-1/2p} \right\}$.
which provides the second term in (4.2) and completes the proof of the Theorem.

For fixed \( j \) the second term in (4.2) is minimized by taking \( t_j = j/J \) (i.e., taking the intervals \( I_j \) of equal length \( 1/J \)). By suitable choice of the sequences \( v_j, u_j \) as initial segments of infinite sequences - see Theorem 3.6 of Niederreiter (1978) - one can realize the (order-of-magnitude) lower bound

\[
(4.12) \quad D_k(v_j, u_j) = O\left( \frac{\log n}{k} \right), \quad \text{for each} \; j. \; (Two \; suitable \; sequences \; are \; the \; Hammersley \; sequence \; and \; the \; Halton \; sequence.) \; Therefore, \; the \; bound \; (4.2) \; assumes \; the \; form
\]

\[
\| \dot{f} - f \|_p \leq C(\log n)^{1/2} + J^{-1/2p},
\]

where \( C \) is a constant.

For \( n = JK + 1 \) and an appropriate allocation of effort between numerical integration and interpolation, we obtain the following result.

**COROLLARY 4.1.** Let \( J_n = n^{p/(p+1)} \) with the \( I_j \) all of length \( 1/J_n \) and let \( K_n = n^{1/(p+1)} \) with the sequences \( v_j, u_j \) chosen to satisfy (4.12). Then

\[
(4.13) \quad \| \dot{f} - f \|_p = O((\log n)^{1/2} n^{-1/(2p+1)}), \quad \text{as} \; n \to \infty.
\]

When \( f \) is continuous the inequality

\[
\| \dot{f} - f \|_p \leq \| \dot{f} - f \|_p,
\]

together with (3.2), (3.7) or (3.8) sometimes, but not always, yields a better estimate than those derived in this section. For example, for \( f \in \text{Lip} \; 1 \; and \; p = 1 \), the rate of convergence in (3.8) is \( O(n^{-1/2}) \), compared to the rate \( O((\log n)^{1/2} n^{-1/3}) \) in (4.13).

However, as \( f \) becomes less smooth, (3.8) deteriorates whereas (4.13) does not. If \( f \in \text{Lip} \; 1/4 \), then the rate in (3.8) is \( O(n^{-1/2}) \), but the rate (4.13) remains \( O((\log n)^{1/2} n^{-1/3}) \). Also in comparison with Theorem 3.1, we note that the bound in Theorem 4.1 depends rather strongly on the distribution of the \( t_{jk} = t_{j-1} + \nu_k (t_j - t_{j-1}) \) within the interval \( I_j \), whereas the bound in Theorem 3.1 exhibits no dependence on the distribution of the \( t_{jk} \).

As in Section 3 we may replace the quasi-random numbers \( u_{jk} \) by independent, uniform \([0,1]\) random variables \( u_{jk} \) and obtain an almost sure rate for the convergence of \( \| \dot{f} - f \|_p \) to zero. The estimator below incorporates an efficient division of labor between numerical integration and interpolation for specified \( n = JK \); this division depends on \( p \).

**PROPOSITION 4.1.** Assume that \( |f| \leq M \) and that \( f \) is of bounded variation on \([0,1]\). Let \( p \in [1,\infty) \) be fixed and for each \( n \) let

\[
(4.14) \quad \bar{f}_n(t) = \frac{1}{K} \sum_{k=1}^{K} q(f \left\{ \frac{k+\nu_k}{n} \right\}, u_{jk}), \quad t \in I_j,
\]
where \( I_\j = \{ \j-1/n, 1/n \}, \) \( \j = \j_n = n^{p/(p+2)}, \) \( K = K_n = n^{2/(p+2)}, \)

where \( v_{jk} = (2k-1)/2K \) for each \( j \) and \( k, \) and where the \( v_{jk} \) are independent, uniform \([0,1]\) random variables. Then almost surely

\[
\|f_n - f\|_P = O((\log n)^{1/4} n^{-1/2} (p+2)) \quad n \to \infty.
\]

Note that the \( v_{jk} \) of \((4.1)\) remain nonrandom and, in fact, are taken to be the minimum discrepancy sequence of length \( K.\)

The proof of Proposition 4.1 is based on a multidimensional analogue of Lemma 3.2, due to Kiefer (1961).

**Lemma 4.1.** Let \( \ang, \ou \) be a sequence of independent random variables, with each \( \ang, \ou \) uniformly distributed in \([0,1]^2.\) Then for each \( \e > 0 \) there is a constant \( C \) (depending on \( \e \) but not on \( K.\) ) such that

\[
P(\max_{1 \leq j, k \leq n} \ang(v_j, u_j) > \e) \leq C n^{-(2-\e)} e^2
\]

for all \( \e > 0.\)

**Proof of Proposition 4.1.** To begin, we observe that

\[
P(\max_{1 \leq j, k \leq n} \ang(v_j, u_j) > \e^{1/2}) \leq J_n \|f_n - f\|_P \|u_j - u\|_2.
\]

If the \( v_{jk} \) are independent, uniform \([0,1]\) random variables independent of the \( u_{jk}, \) then since \( u_j \) is the minimum discrepancy sequence of

length \( K_n, \) we have

\[
P(\max_{1 \leq j, k \leq n} \ang(v_j, u_j) > \e^{1/2}) \leq O(n^{1/2} n^{-1/2} (p+2)) = O(\log n)^{1/4} n^{-1/2} (p+2),
\]

by conditioning on \( u_j.\) Now choose \( \e > 0 \) sufficiently small that \( \e < 1-p/(p+2), \) and let

\[
d_n = (\log n)^{1/4} n^{-1/2}.
\]

Applying \((4.16) - (4.18)) we infer that

\[
P(\max_{1 \leq j, k \leq n} \ang(v_j, u_j) > \e^{1/2}) \leq O(d_n) = O(n^{-1/2} (p+2)) = O(d_n),
\]

Therefore by the Borel-Cantelli lemma,

\[
\max_{1 \leq j, k \leq n} \ang(v_j, u_j) > \e^{1/2} = O(d_n),
\]

almost surely. The proof now follows by combining \((4.2)\) and \((4.19),\)

with the stated choice of \( J_n.\)

For \( p = 1, \) the rate of convergence in \((4.15)\) is \((\log n)^{1/4} n^{-1/2}, \)

which, while slower than that of \((3.11)\) for \( f \in \text{Lip } 1, \) exceeds the latter rate for less smooth \( f \) and even applies to discontinuous \( f \) of bounded variation.
5. Alternative Methods of Estimation in $L^2$. For $L^2$-estimation of continuous $f$, there are other ways to use the data \( \{(t_{jk}, x_{jk})\} \) to estimate $f$. Two such methods are developed in this section. The first involves an estimator of rather different form than the $f$ given by (2.3), and for it we obtain an error bound less than that in (3.1) but of the same order of magnitude. The second method uses an adaptive choice of the $u_{jk}$ for each $j$ and achieves a dramatically improved rate of convergence: $n^{-1}$ for all $f$ satisfying any Lipschitz condition.

The methods of this section are based on the observation that the value of $x_{jk}$ determines whether the point $(t_{jk}, 2b_{jk} + B)$ is below, on, or above the graph of $f$, and that given two such points on opposite sides of the graph, the line segment joining them intersects the graph of $f$ (by continuity of $f$).

Since discrepancy does not enter the analysis, it is convenient to replace the $u_{jk}$ by points $x_{jk} \in (-B, B)$, where $B$ is the bound on $|f|$. The data, therefore, is represented as

\[
(5.1a) \quad \{(t_{jk}, x_{jk}, b_{jk}); 1 \leq j \leq J, 1 \leq k \leq K\},
\]

where the $t_{jk}$ satisfy (2.3b), the $x_{jk}$ are in $[-B, B]$, and

\[
(5.1b) \quad b_{jk} = \text{sgn} \left( f(t_{jk}) - x_{jk} \right).
\]

Note that whereas in Section 2 the $u_{jk}$ were not treated as data (since their values are not necessary to compute the estimator $\hat{f}$, although they do appear in the bound (3.1)) the $x_{jk}$ now must be part of the data since their values are needed to calculate the estimator $\hat{f}$ defined in (5.2) below. We assume, still, that there are intervals $I_j = [t_{j-1}, t_j]$ satisfying $t_{jk} \in I_j$ for each $j$ and $k$ and that the estimator $\hat{f}$ is to be a step function satisfying (2.4) whose value on $I_j$ depends only on $\{(t_{jk}, x_{jk}, b_{jk}); 1 \leq k \leq K\}$.

For the first estimator the $x_{jk}$ are fixed in advance, i.e., the estimator is nonadaptive. Its value on $I_j$ is determined as follows: Let $-B = x_{j,0} \leq x_{j,1} \leq \cdots \leq x_{j,K} \leq x_{j,K+1} = B$ be the order statistics of the points $-B, x_{j,1}, \ldots, x_{j,K}$, $B$.

Let $\sigma$ be that permutation of $\{1, \ldots, K\}$ for which $x_{j,\sigma(1)} = x_{j,1}, \ldots, x_{j,\sigma(K)} = x_{j,K}$, and $\{t_{j}, x_{j,\sigma(i)}\}$ in that order (these are merely the points $(t_{j-1}, -B), (t_{j,1}, x_{j,1}), \ldots, (t_{j,K}, x_{j,K}), (t_j, B)$ with the second coordinates in increasing order). Since $f$ is continuous at least one segment of $\hat{f}$ intersects the graph of $f$ and which segments do so can be determined from the data (5.1). Choose any such segment, say that with endpoints $\{t_{j,\sigma(1)}, x_{j,\sigma(1)}\}$ and $\{t_{j,\sigma(i+1)}, x_{j,\sigma(i+1)}\}$, and define...
(5.2) \[ \hat{e}(t) = \frac{1}{2} x_{3,j}(t+1) + x_{3,j}(t), \quad t \in J_j. \]

The following result provides an error bound for this estimator.

**Theorem 5.1.** Assume that \( f \) is continuous and that \( |f| \leq b \) on \([0,1]\) and let \( \hat{e} \) be given by (5.2). Then

\[ \|\hat{e} - f\|_\infty \leq \frac{1}{2} \max_{J \subseteq J} \max_{k, l} \left( |x_{3,j}(k) - x_{3,j}(k-1)| + |x_{3,j}(l) - x_{3,j}(l-1)| + w(f, t_j - t_{j-1}) \right) \]

(Recall that \( w(f, \cdot) \) is the modulus of continuity of \( f \)).

**Proof.** Fix \( j \). If \( \hat{e}(t), t \in J_j \), is given by (5.2), then in the interval with endpoints \( t_{j,i}(\cdot) \) and \( t_{j,i}(\cdot+1) \), there is \( t \) such that \( f(t) \) lies on \( \Gamma \). Hence for \( t \in J_j \),

\[ |\hat{e}(t) - f(t)| \leq |\hat{e}(t) - \hat{e}(t)| + |\hat{e}(t) - f(t)| \]

\[ \leq \frac{1}{2} |x_{3,j}(t+1) - x_{3,j}(t)| + w(f, t_j - t_{j-1}), \]

and (5.3) follows immediately.

With \( J \) and \( K \) fixed the right-hand side of (5.3) is minimized by choosing \( t_j = \alpha J \) and the \( x_{3,j} \) such that \( x_{3,j}(\cdot) = x_{3,j}(\cdot-1) = ZB(K+1)^{-1}. \) This transforms (5.3) to

\[ \|\hat{e} - f\|_\infty \leq \frac{1}{2} \max_{J \subseteq J} \max_{k, l} \left( |x_{3,j}(k) - x_{3,j}(k+1)| \right) + w(f, t_j - t_{j-1}), \]

which slightly improves (3.6) for finite \( K \) and is equivalent in order as \( K^m \). If the \( x_{3,j} \) are increasing in \( k \) for each \( j \), then the computational and storage requirements for (2.5) and (5.2) are comparable. However, whereas (2.5) can be updated recursively as \( K \) increases, (5.2) cannot.

**Corollary 5.1.** If \( f \) is Lipschitz \( \gamma \) satisfies the assumptions of Theorem 5.1 and \( \hat{e} \) is given by (5.2) with \( J = J_n = n^{1/(r+1)} \) and \( t_j = j/j \) and with \( x_{3,j} = x_{3,j}(k) \) for each \( j \) and \( k \), then

\[ \|\hat{e} - f\|_\infty = o(n^{-1/(r+1)}), \]

Using the easily established inequality

\[ \max_{J \subseteq J} \max_{k, l} \left( |x_{3,j}(k) - x_{3,j}(k+1)| \right) \leq 4B(K+1)^{-1}, \]

where \( x_{3,j} = ZB(K+1)^{-1}, \) one can obtain an almost sure convergence rate for the case where the \( x_{3,j} \) are replaced by independent, uniform \([-B,B] \) random variables \( x_{3,j} \); the rate is that of (3.11). We omit further details.

Although seemingly unrelated, the methods yielding the estimators (2.5) and (5.2) can be viewed as two approaches to the same numerical integration problem. For \( f \in [-B,B] \), recall that \( \gamma = \int_{-B}^B \gamma(y) dy \). The method of Sections 2 and 3 approximates the integral - within the restrictions we have imposed - as an average of q-
values. However, for fixed y, the function u = q(y,u) assumes only the known values -B, 0 and B so that its integral can be estimated from an estimate of its single point of discontinuity, namely 

\[ u(y) = (B - y)/2B. \]

The estimator \( \hat{f} \) in (5.3) estimates this point of discontinuity from the data (5.1).

Using the nonadaptive estimator \( \hat{f} \) of (5.2) one can "pin down" the value of \( f \) somewhere in the interval \( I_j \) only to within 

\[ \max(x_j, \alpha_j, \beta_j, x_{j-1}) \leq 2B^{-1}. \]

By choosing the \( x_{jk} \) adaptively, one can do much better and can obtain rates of convergence \( \| \hat{f} - f \|_w = O((\log n)^{-1}) \) for all \( f \) satisfying any Lipschitz condition. We now describe the algorithm for constructing \( \hat{f} \), which is based on the data (5.1), except that now the \( x_{jk} \) will be determined adaptively - but recursively - for each \( j \). Assume that \( |f| \leq B \) on \([0,1] \). The value of \( \hat{f} \) on \( I_j \) is constructed as follows.

**Step 1 (Initialization).** Note that the points \( (t_{j-1}^-, x_{j-1}^-) \) and \( (t_{j-1}^+, x_{j-1}^+) \) are above and below the graph of \( f \), respectively. Set 

\[ (t_0^-, x_0^-) = (t_{j-1}^-, x_{j-1}^-), \quad (t_0^+, x_0^+) = (t_{j-1}^+, x_{j-1}^+). \]

**Step 2 (Iteration).** The \( k \)th step of the iteration is entered with two points \( (t_{k-1}^-, x_{k-1}^-) \) and \( (t_{k-1}^+, x_{k-1}^+) \), from among the points 

\[ (t_{j-1}^-, x_{j-1}^-), (t_{j-1}^+, x_{j-1}^+), \ldots, (t_{j-k+1}^-, x_{j-k+1}^-), (t_{j-k}^+, x_{j-k}^+), \]

such that \( (t_{k-1}^-, x_{k-1}^-) \) is above the graph of \( f \), \( (t_{k-1}^+, x_{k-1}^+) \) is below the graph of \( f \), and \[ |x_{k-1}^- - x_{k-1}^+| = B^{-k/2}. \] Then set \( x_{jk} = \frac{1}{2} (x_{k-1}^- + x_{k-1}^+) \) and calculate \( b_{jk} = \min(f(t_{jk}) - x_{jk}) \).

a) If \( b_{jk} = 0 \), then \( f(t_{jk}) = x_{jk} \) proceed to the termination step.

b) If \( b_{jk} = 1 \), then \( f(t_{jk}, x_{jk}) \) is below the graph of \( f \). Let 

\[ (t_{k}^-, x_k^-) = (t_{k-1}^-, x_{k-1}^-) \] and \( (t_k^+, x_k^+) = (t_{jk}, x_{jk}) \), and proceed to the next iteration.

c) If \( b_{jk} = -1 \), then \( f(t_{jk}, x_{jk}) \) is above the graph of \( f \). Let 

\[ (t_{k}^-, x_k^-) = (t_{jk}, x_{jk}) \] and \( (t_k^+, x_k^+) = (t_{k-1}^+, x_{k-1}^+) \), and proceed to the next iteration.

Note that for either b) or c) we have \[ |x_k^- - x_k^+| = \frac{1}{2} |x_{k-1}^- - x_{k-1}^+|. \]

**Step 3 (Termination).** There are two possibilities.

a) If there is \( k \) such that \( f(t_{jk}) = x_{jk} \), set 

\[ f(t) = x_{jk}, \quad t \in I_j. \]

b) Otherwise, the procedure terminates with evaluation of \( b_{jk} \) and yields points \( (t_k^-, x_k^-) \) above the graph of \( f \) and \( (t_k^+, x_k^+) \) below the graph of \( f \) such that \[ |x_k^- - x_k^+| = B^{-k+1}. \] In this case set 

\[ f(t) = \frac{1}{2} (x_k^- + x_k^+), \quad t \in I_j. \]

Note that the \( t_{jk} \) are not chosen adaptively, but can be specified in advance. Furthermore, the algorithm is recursive in that only the current values of \( (t_k^-, x_k^-) \) and \( (t_k^+, x_k^+) \) need to be stored in order to determine either \( x_{jk}, b_{jk} \) or the value of the estimator.

The resulting error bound dramatically improves those for the nonadaptive estimators (2.5) and (5.2).

**Theorem 5.2.** Assume that \( f \) is continuous and that \( |f| \leq B \) on \([0,1] \) and let \( f \) be given by (5.6). Then

\[ \|f - \hat{f}\|_w \leq B^{-k+1} + \max_{1 \leq j \leq k} \left\{ \frac{1}{2} |f(t_{jk}) - x_{jk}| \right\}. \]
Proof. Let \( j \) be fixed and consider the two possible forms of termination separately.

a) If \( f(t_{j+1}^{1}) = s_{j} \) and \((5.6)\) holds, then for \( t \in I_{j} \),
\[
\| \hat{\varepsilon}(t) - f(t) \| = \| f(t_{j+1}^{1}) - f(t) \| \leq w(f; t_{j+1}^{1} - t_{j+1}) \cdot \frac{1}{\sqrt{n}}.
\]

b) If \( f \) is given by \((5.6)\) and continuity of \( f \) there is \( \hat{\varepsilon} \) within the interval with endpoints \( t_{j}^{k} \) and \( t_{k}^{j} \) such that \( f(t_{j}) \) is on the line segment joining \( (t_{k}^{j}, x_{k}^{j}) \) and \( (t_{k}^{j}, x_{k}^{j}) \). Hence, for \( t \in I_{j} \),
\[
\| \hat{\varepsilon}(t) - f(t) \| \leq \frac{1}{2} | x_{k}^{j} - x_{k}^{j} | + | f(t_{j}) - f(t) | \leq \frac{1}{2} | x_{k}^{j} - x_{k}^{j} | + w(f; t_{j} - t_{j+1}) \cdot \frac{1}{\sqrt{n}}.
\]

Consequently, \((5.7)\) holds.

For fixed \( j \), the right-hand side of \((5.7)\) is minimized for \( t_{j} = 1/\sqrt{n} \), yielding the bound
\[(5.8)\]
\[
\| \hat{\varepsilon} - f \| \leq \frac{1}{2} | x_{k}^{j} - x_{k}^{j} | + w(f; t_{j}) \cdot \frac{1}{\sqrt{n}}.
\]

Corollary 5.2. Assume that \( f \in Lip_{\varepsilon} \) and that \( \| f \| \leq B \) on \([0,1]^{d} \). Let \( f \) be given by \((5.6)\) with \( K = n^{j} \), satisfying \( \frac{1}{2} d \sqrt{n} \leq n \). Then \( t_{j} = 1/\sqrt{n} \) and \( t_{j} = \frac{1}{2} \sqrt{n} \) for each \( j \).

\[(5.9)\]
\[
\| \hat{\varepsilon} - f \| = O(\log n)^{1/2} \cdot \frac{1}{\sqrt{n}}.
\]

Proof. From \((5.8)\), there is a constant \( C \) such that
\[
\| \hat{\varepsilon} - f \| \leq C \left( \frac{1}{\sqrt{n}} \right) \cdot \frac{1}{\sqrt{n}} \cdot \frac{\log n}{n}.
\]

The algorithm leading to \((5.6)\) can be improved in practice, although not in the worst case (i.e., the bound \((5.7)\) is not improved), by the following device. If at any iteration the points \( (t_{k}^{j}, x_{k}^{j}) \) above the graph of \( f \) and \( (t_{k}^{j}, x_{k}^{j}) \) below satisfy \( x_{k}^{j} < x_{k}^{j} \), which is easily checked, then somewhere in the interval with endpoints \( t_{k}^{j}, t_{k}^{j} \), \( f \) must acquire the value \( \frac{1}{2} (x_{k}^{j} + x_{k}^{j}) \). If one takes \( f(t_{j}) \) to be this value, then the first term on the right-hand side of \((5.7)\) is unnecessary and one has \( \| \hat{\varepsilon} - f \| \leq w(f; t_{j}) \cdot \frac{1}{\sqrt{n}} \) on \( I_{j} \).

6. Complements. In this section we sketch an extension of Theorems 1.1 to the case of functions \( f \) defined on \([0,1]^{d} \) for some \( d \geq 2 \). In addition, we include a few comments concerning our method of reconstruction.

We first consider reconstruction of functions on \([0,1]^{d} \). Let \( n \geq 2 \).

The estimator \( \hat{f} \) is constructed in the following manner. Partition \([0,1]^{d} \) into intervals \( t_{j}^{0} \) as in Section 1 and let the \( t_{j}^{0} \) satisfy \((2.3)\). Suppose that for each choice of \( j = (j_{1}, \ldots, j_{d}) \), where \( 0 \leq j_{p} \leq d \) for each \( j_{p} \), there are \( n^{d} \) numbers \( \{w_{j_{1}, \ldots, j_{d}} \} \), \( 1 \leq k \leq n \), in \([0,1]^{d} \). By analogy with \((2.5)\), we introduce the estimator
\[(6.1)\]
\[
\hat{f}(t) = \frac{1}{n^{d}} \sum_{j=1}^{n} w_{j_{1}, \ldots, j_{d}} f(t_{j_{1}}, \ldots, t_{j_{d}}), \quad t \in I_{1} \times \cdots \times I_{d}.
\]

The work (i.e., number of function evaluations) required to calculate \( f \) is \( n^{d} \).
An essentially verbatim repetition of the proof of Theorem 3.1 yields

**Theorem 6.1.** Assume that \( |x| \leq b \) on \([0,1]^d\) and that \( f \) is defined by (6.1). Then

\[
\| \hat{f} - f \|_m \leq 2b \max_{k \leq 3} \| D^s_k (u_{j,k}) \| + \max_{k \leq 3} \| f(x_{j,k}) \|,
\]

where \( D^s_k (u_{j,k}) \) is the discrepancy (in \([0,1]^d\)) of the numbers \( \{u_{j,k} : 1 \leq k \leq n\} \), and where \( u(f; \mathcal{A}) \) is the oscillation function of \( f \), given by

\[
u(f; \mathcal{A}) = \sup \{ |f(t) - f(s)| : t, s \in \mathcal{A}\}.
\]

By suitable choice of the \( u_{j,k} \) - see the comments following Theorem 3.1 - and for equally spaced \( \tau_j \), it is easily verified that if \( f \) satisfies

\[
|f(t) - f(s)| \leq \| \xi - \eta \| Y
\]

for some \( \gamma \in (0,1) \), where \( \| \xi - \eta \| \) is the Euclidean norm of \( \xi - \eta \), then with an optimal division of labor between integration and interpolation,

\[
\| \hat{f} - f \|_m = O(n^{-\gamma/(d+\gamma)}),
\]

where \( n = (\lambda x)^d \) is the work required to calculate \( f \) using (6.1).

The results in Section 4 on \( L^p \) convergence extend analogously.

We conclude with some comments concerning our method for reconstructing a function. First, our bounds for errors of the form \( \| \hat{f} - f \|_m \) are insensitive to the distribution of the points \( t_{j,k} \) in each interval \( \tau_j \). It is possible that in some cases one could have, if desired, \( t_{j,1} = \ldots = t_{j,k} \). (However, in many applications this will be impossible because of the sequential nature of the data collection procedure.) Being able to evaluate \( q(f(t_{j,k}), u_{j,k}) \) for \( K \) values of \( u_{j,k} \) in effect permits discritization of the function \( f \) at \( t_{j,k} \). If the \( u_{j,k} \) are increasing in \( k \), this is precisely what the estimator (5.2) accomplishes. One cannot improve the bound (5.3) by taking the \( t_{j,k} \) to be equal for fixed \( j \), nor can any other of our \( L^p \) bounds be similarly improved. Regarded in this context, our estimator (2.5) is as effective as the estimators obtained by discretizing \( f \) at one point in each interval \( \tau_j \).

Here are two final points. In practice the bound \( B \) on \( f \) may not be known; our estimators then serve simply to estimate the truncated function \( f_B \) defined by

\[
f_B(t) = \begin{cases} f(t) & \text{if } f(t) > B, \\ 0 & \text{if } f(t) \leq B. \end{cases}
\]

Also, the restriction to \([0,1]\) as the domain of \( f \) is essential; one can replace \([0,1]\) by any finite interval \([a,b]\), although the constants in some of the bounds will be multiplied by \( b-a \). Since our estimators are local on the intervals \( \tau_j \), estimation over, for example, \([0,m] \) is even possible.
In effect, our procedure estimates f based on observations through a window of the form \([0,1] \times [-R,R]\), which can be replaced by any other rectangular window.

Acknowledgement. We are indebted to our colleague Richard H. Byrd for some discussions stimulating the development of Section 5.

References


---

**Unclassified**

**SECURITY CLASSIFICATION OF THIS PAGE**

<table>
<thead>
<tr>
<th>1. REPORT NUMBER</th>
<th>2. GOVT ACCESSION NO.</th>
<th>3. RECIPIENT CATALOG NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4A-2604</td>
<td></td>
</tr>
</tbody>
</table>

---

**REPORT DOCUMENTATION PAGE**

<table>
<thead>
<tr>
<th>4. TITLE</th>
<th>5. TYPE OF REPORT &amp; PERIOD COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION F FROM A FINITE SEQUENCE [SN(F(t) + x)]</td>
<td>Technical Report</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>7. AUTHOR(s)</th>
<th>8. CONTRACT OR GRANT NUMBER(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan F. Karr and Robert J. Serfling</td>
<td>OMR No. NO001-79-C-0801</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>9. PERFORMING ORGANIZATION NAME AND ADDRESS</th>
<th>10. PROGRAM ELEMENT, PROJECT, TASK AREA &amp; WORK UNIT NUMBERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Mathematical Sciences</td>
<td></td>
</tr>
<tr>
<td>The Johns Hopkins University</td>
<td></td>
</tr>
<tr>
<td>Baltimore, Maryland 21218</td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>11. CONTROLLING OFFICE NAME &amp; ADDRESS</th>
<th>12. REPORT DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office of Naval Research</td>
<td>May, 1981</td>
</tr>
<tr>
<td>Statistics and Probability Program</td>
<td></td>
</tr>
<tr>
<td>Arlington, Virginia 22217</td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>13. NUMBER OF PAGES</th>
<th>15. SECURITY CLASS (of this report)</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>Unclassified</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>16. DISTRIBUTION STATEMENT (of this report)</th>
<th>15a. DECLASSIFICATION/DEMERITATION SCHEDULE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved for public release; distribution unlimited.</td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>18. SUPPLEMENTARY NOTES</th>
</tr>
</thead>
</table>

---

<table>
<thead>
<tr>
<th>19. KEY WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonparametric signal identification; approximation; convergence rates.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>22. ABSTRACT</th>
<th>25. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider reconstructing a function f(t), 0 ≤ t ≤ 1, from knowledge only of ((f(t_i) + x_i), 1 ≤ i ≤ n), where (x_i = \text{sgn}(f(t_i)) \cdot x_i), (1 ≤ i ≤ n), and the (x_i) are additive &quot;corruptions.&quot; Without these (x_i), (f) could not be reconstructed. However, if (f) is continuous and (f) is random uniform (x), Marry and Cambanis (1980, 1981) show that (f) can be consistently estimated almost surely in norm square as (n^{-1}). In the present treatment the approximation of (f(t_i)) is identified as a numerical integration problem rather than a statistical problem. We obtain simple bounds on the error of estimation, allow arbitrary (random or deterministic) noise (x_i), and deal with the case of discontinuous (f). The bounds yield substantially improved convergence rates for (x_i) a quasi-random rather than random sequence.</td>
<td></td>
</tr>
</tbody>
</table>

---