The Annals of Applied Probability
1997, Vol. 7, No. 3, 648–665

APPROXIMATING A LINE THROWN AT RANDOM ONTO A GRID

BY PETER HALL AND MARC RAIMONDO

Australian National University and CSIRO, and Australian National University

Throw a straight line at random into a plane, within which is inscribed a square grid. Color black each grid vertex that lies above the line, and white each vertex below it. Now remove the line, and attempt to reconstruct it from the pattern of vertex colors in an $m \times m$ section of the grid. We show that for all $\varepsilon > 0$, the line can be approximated to within order $m^{-1}(\log m)(\log\log m)^{1+\varepsilon}$, with probability 1, and that there is no deterministic subsequence along which the best achievable rate is better than $(m \log m)^{-1}(\log\log m)^{-1-\varepsilon}$ with positive probability. Both these results fail if $\varepsilon = 0$. More generally, we provide a complete characterization of almost sure rates of approximation, in terms of the convergence or divergence of an infinite series. Applying these results, we develop near-optimal local linear approximations to general smooth boundaries. We address the case where vertex color is a shade of gray, varying in the continuum and subject to stochastic error.

1. Introduction. Suppose a line has been placed randomly into a plane, within which is inscribed a regular grid. For definiteness we consider a square grid, although our results hold also for other regular grids. Color black those grid vertices that lie above the line, and white those below it. Now remove the line, and attempt to reconstruct it from the pattern of vertex colors within an $m \times m$ square in the grid. We show that, with probability 1 with respect to the distribution of the line, the position of the line may be determined to within "approximately" $O(m^{-1})$. More concisely, if L is a positive, slowly varying function, then the line can be estimated with accuracy $m^{-1}(\log m)/L(\log m)$, with probability 1, if and only if

$$\sum_{n=1}^{\infty} n^{-1}L(n) < \infty; \tag{1.1}$$

and the best convergence rate along the most optimistically chosen deterministic subsequence of values of m is, with probability 1, inferior to $(m \log m)^{-1}L(\log m)$ if and only if (1.1) is violated. (The definition of randomness for the line can be very general. We require only that the distribution of

Received May 1996; revised March 1997.
AMS 1991 *subject classifications*. Primary 60G35; secondary 62G20.
Key words and phrases. Digital image, gradient, grid, irrational number, local linear, rational number, slope, vertex.

its slope conditional on its intercept be absolutely continuous, for each intercept in the support of the distribution of that quantity.)

We apply these results to produce accurate, local linear approximations to twice-differentiable boundaries. That work is for general, monochromatic images recorded in the continuum, with a boundary represented by a discontinuity in a smoothly varying range of gray colors, observed with noise. Depending on the number of moments assumed of the noise distribution, the order of approximation provided by our local linear smoother can be very close to the optimum for an idealized, black-or-white, noiseless image.

Aspects of this problem have received significant attention in literature on the theory of digital imaging. In particular, optimal estimation of boundaries from vertex data has been addressed at length by Korostelev and Tsybakov (1993), although largely without discussion of the case of a regular grid. Properties of boundary approximations computed from data on regular-grid vertices are quite different from those where vertices arise randomly, for example when they are points of a Poisson process in the plane. In particular, if we wish to use vertex colors to approximate a line that is parallel to one of the grid axes then, unless it passes through a vertex, the accuracy of our approximation is limited solely by the edge width of the grid, and does not depend on the amount of data observed. This means that results on minimax efficiency, presented by Korostelev and Tsybakov (1993) for a random distribution of vertices, do not extend to gridded data.

After a little consideration it is clear that estimation of a straight line with any *rational* slope is fraught with the same difficulties that beset approximation of a line that is parallel to one or other of the coordinate axes. For a line with rational slope, the best order of approximation, using the color pattern in the whole plane, is $O(1)$ unless the line intersects a vertex. If the line is chosen at random, however, according to the definition of "random" suggested in the first paragraph, then its slope is irrational with probability 1, and so these difficulties do not arise. Our technical arguments are founded on methods from stochastic number theory, describing the accuracy of rational approximations to randomly chosen irrationals.

The dichotomy of radically different rates of approximation for lines with rational and irrational slopes is apparent in empirical studies. If one determines numerically the region $\mathscr{R}$, within which a line must lie in order to produce a given color pattern in a certain finite set $\mathscr{S}$ of vertices, and then gradually alters the line's slope while fixing $\mathscr{S}$, then the width of $\mathscr{R}$ fluctuates erratically with slope. This reflects the fact that both the rationals and irrationals are dense in the reals, and also the variability of the convergence rate among irrational slopes.

Early work on methods for approximating boundaries using vertex color patterns includes that of Kulpa (1977) and Profitt and Rosen (1979). Other techniques, and their theoretical properties, have been discussed by (among others) Dorst and Smeulders (1984, 1986, 1987), Hung (1985), McIlroy (1985) and Koplowitz and Bruckstein (1989). Brady and Asada (1984) and Landau (1987) gave algorithms for approximating boundaries. See also Freeman

(1970) and Groen and Verbeek (1978). The "Freeman code" provides a particularly effective technique for tracking a boundary past a sequence of colored vertices. See Worring and Smeulders (1995) for recent applications.

Our basic results about orders of approximation to random lines will be presented in Section 2. Their ramifications will be discussed in Section 3. These include the accuracy of local linear approximations to general boundaries, and the effect of noise. Technical arguments will be included together in Section 4.

2. Approximating a straight line.

2.1. *Approximation to a linear boundary in gridded data.* A straight line $\mathscr{L}$ placed into a plane divides it into two halves, of which that above the line we shall imagine to be colored black, and that below to be colored white. (It is assumed that the line is not vertical.) Consider a square grid in the plane whose vertices are at integer coordinate pairs, relative to a Cartesian system. A vertex of the graph will be said to be black or white, depending on the half of the plane in which it lies. Only lines with irrational slope are relevant to our analysis, and of course such a line can pass through at most one vertex. An arbitrary color (black, white or "colorless") may be assigned to this vertex without affecting our results.

Let $\mathscr{S} = \mathscr{S}(m)$ denote the set of all vertices within the square $[-\frac{1}{2}m, \frac{1}{2}m] \times [-\frac{1}{2}m, \frac{1}{2}m]$, where m is an integer. From perfect knowledge of the colors of vertices within $\mathscr{S}$ we wish to approximate the position of $\mathscr{L}$. Given a specific black and white vertex pattern $\mathscr{P} = \mathscr{P}(\mathscr{L})$ within $\mathscr{S}$, produced by $\mathscr{L}$, let $\hat{\mathscr{L}}(\mathscr{P})$ denote an approximation to $\mathscr{L}$. (We suppress the dependence of $\mathscr{P}$ and $\hat{\mathscr{L}}$ on m.) For example, $\hat{\mathscr{L}}$ might be constructed by a least-squares method, of which there are several.

Next we introduce a measure of the distance between $\mathscr{L}$ and $\hat{\mathscr{L}}$ within $\mathscr{S}$. If $\mathscr{L}$ and $\hat{\mathscr{L}}$ intersect, let $\tilde{\mathscr{L}}$ denote the line that passes through the point of intersection and bisects the angle between $\mathscr{L}$ and $\hat{\mathscr{L}}$. If $\mathscr{L}$ and $\hat{\mathscr{L}}$ are parallel, let $\tilde{\mathscr{L}}$ be the line that is equidistant from them. Let $\mathscr{T}(\mathscr{L}, \hat{\mathscr{L}}, \mathscr{S})$ be the set of line segments that join $\mathscr{L}$ and $\hat{\mathscr{L}}$, are perpendicular to $\tilde{\mathscr{L}}$, and are contained within $\mathscr{S}$. Define $D(\mathscr{L}, \hat{\mathscr{L}})$, the Hausdorff distance between those parts of $\mathscr{L}$ and $\hat{\mathscr{L}}$ that lie within $\mathscr{S}$, to be the supremum of the lengths of elements of $\mathscr{T}(\mathscr{L}, \hat{\mathscr{L}}, \mathscr{S})$.

2.2. *Approximation to irrational numbers by continued fractions.* A real number u that is not an integer may be uniquely expressed as a continued fraction,

$$u = [a_0; a_1, a_2, \dots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}},$$

where a_0 is an integer and $a_1, a_2, \dots$ are strictly positive integers, called the *partial denominators* of u. The continued fraction expansion terminates if and only if u is rational. Up to the termination point (in the case of a rational), or for all n (if u is irrational), the convergents of u are the numbers

$$\frac{p_0}{q_0} = a_0, \qquad \frac{p_1}{q_1} = a_0 + \frac{1}{a_1}, \qquad \frac{p_2}{q_2} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2}}, \ldots,$$

where p_n and q_n are positive and relatively prime integers. The reader is referred to Khintchine (1963) for a discussion of continued fractions and their properties.

We say that u is *badly approximable* (or BA, for short), if $\sup_n a_n(u) < \infty$. The set of all BA numbers in $(0, 1)$ has cardinality equal to that of the continuum [e.g., Schmidt (1980), page 23], but is of measure zero [e.g., Khintchine (1963), page 69]. All quadratic irrationals are BA.

2.3. *Main results.* Let $\mathscr{U}(\mathscr{P})$ denote the set of all lines $\hat{\mathscr{L}}$ that produce the observed pattern $\mathscr{P}$ of vertex colors within $\mathscr{S}$. Each $\hat{\mathscr{L}}$ may be considered an approximation to $\mathscr{L}$. Let P denote probability measure with respect to any given random distribution of $\mathscr{L}$, such that the distribution of slope conditional on intercept is absolutely continuous.

THEOREM 2.1. (a) *For any positive function L that is slowly varying at infinity, the following three assertions are equivalent*:

$$P\left\{\sup_{\hat{\mathscr{L}} \in \mathscr{U}(\mathscr{P})} D(\mathscr{L}, \hat{\mathscr{L}}) = O\left(\frac{\log m}{mL(\log m)}\right)\right\} = 1,$$

$$P\left\{\liminf_{m\to\infty} \frac{m \log m}{L(\log m)} \sup_{\hat{\mathscr{L}} \in \mathscr{U}(\mathscr{P})} D(\mathscr{L}, \hat{\mathscr{L}}) > 0\right\} = 1,$$

$$\sum_{n=1}^{\infty} n^{-1} L(n) < \infty.$$

(b) *If the gradient of $\mathscr{L}$ is a BA irrational then*

$$\sup_{\hat{\mathscr{L}} \in \mathscr{U}(\mathscr{P})} D(\mathscr{L}, \hat{\mathscr{L}}) = O(m^{-1}), \tag{2.1}$$

$$\liminf_{m\to\infty} m \sup_{\hat{\mathscr{L}} \in \mathscr{U}(\mathscr{P})} D(\mathscr{L}, \hat{\mathscr{L}}) > 0. \tag{2.2}$$

Furthermore, both (2.1) *and* (2.2) *fail if the gradient is not a BA irrational.*

REMARK 2.1. Theorem 2.1 might be interpreted as stating that "the best achievable rate of approximation to $\mathscr{L}$ by $\hat{\mathscr{L}}$ is $O(m^{-1})$." This is exactly true if and only if the gradient of $\mathscr{L}$ is a BA irrational. It is approximately true if the gradient is chosen at random, in any continuum sense, from the set of all irrational numbers. There, with probability 1 the best achievable rate along the best subsequences is no better than and no worse than

$$(m \log m)^{-1}(\log\log m)^{-1-\varepsilon} \quad \text{and} \quad (m \log m)^{-1}(\log\log m)^{-1+\varepsilon},$$

respectively, for all $\varepsilon > 0$; and the best achievable rate along the worst subsequences is no better than and no worse than

$$m^{-1}(\log m)(\log\log m)^{1-\varepsilon} \quad \text{and} \quad m^{-1}(\log m)(\log\log m)^{1+\varepsilon},$$

respectively, for all $\varepsilon > 0$.

REMARK 2.2 (Nonsquare grids). The vast majority of grids used in practice are square, but the Leitz texture analyzer (for which the mathematical theory was developed by J. P. Serra) employs a hexagonal grid. Our results hold without change for any regular grid (such as a hexagonal one) all points of which are represented by a finite number of square subgrids.

3. Local linear approximation to general boundaries.

3.1. *Model for a boundary on a fine grid.* Given a Cartesian coordinate system in the plane, construct a square grid with vertices at pairs of integer multiples of n^{-1}, where $n \geq 1$ is an integer. Represent a boundary, $\mathscr{C}$, in the plane by an equation $y = g(x)$, where g is a continuous function. Color black or white those parts of the plane that lie above or below the boundary, respectively. A vertex assumes the color of that part of the plane in which it is situated. We are interested in the accuracy with which we can approximate $\mathscr{C}$ from information on vertex colors, perhaps observed with noise, as n increases.

3.2. *Estimating a curved boundary on a fine grid.* Suppose the function g defining $\mathscr{C}$ is differentiable and g' enjoys a Lipschitz condition of order $\gamma - 1$, where $1 < \gamma \leq 2$. It may be proved that, for each $\varepsilon > 0$, $\mathscr{C}$ can be estimated at rate

$$n^{-2\gamma/(\gamma+1)}\{(\log n)(\log\log n)^{1+\varepsilon}\}^{\gamma/(\gamma+1)}$$

by using a local linear approximation within a window. This result is stated for approximation at a random point X and is valid with probability 1 with respect to any continuous distribution of X.

3.3. *Estimation of $\mathscr{C}$ in the presence of noise.* The noiseless data discussed above may be written in the form $Y(i/n, j/n) = I\{j/n \leq g(i/n)\}$, where $I(\cdot)$ is an indicator function, $Y(i/n, j/n)$ denotes the color of the vertex at $(i/n, j/n)$ (white is represented by 1 and black by 0), and the equation $y = g(x)$ represents the boundary $\mathscr{C}$. In practice, due to a combination of systematic and stochastic errors, the color of each vertex may be more appropriately represented by a number between $-\infty$ and ∞. In particular, we may write

$$Y(i/n, j/n) = f(i/n, j/n) + \varepsilon_{ij},$$

where $f(\cdot, \cdot)$ is a function with a fault-type discontinuity along the curve $y = g(x)$ and the independent and identically distributed stochastic errors ε_{ij} have zero mean and are independent of the random point X at which we develop the approximation to $\mathscr{C}$.

It will be assumed that f admits the representation

$$f(x, y) = f_1(x, y) + f_2(x, y)I\{y \le g(x)\},$$

where f_1 and f_2 each have two uniformly bounded derivatives of all types and f_2 is bounded away from zero. We call this condition (C_f). Of g we suppose that it has a derivative which satisfies a Lipschitz condition of order 1 in the interval $[0, 1]$, and that constants $-\infty < C_1 < C_2 < \infty$ are known with the property that $C_1 < g < C_2$ in $[0, 1]$. These assumptions will be referred to collectively as condition (C_g).

We shall introduce a local linear estimator of g which, when (C_f) and (C_g) hold, comes close to achieving the convergence rate described in Section 3.2 in the no-noise case. For example, the convergence rate is $n^{-(4/3)+\delta}$, for any given $\delta > 0$, if the distribution of ε_{ij} has sufficiently many finite moments. This rate is available with probability 1 with respect to the distributions of the errors ε_{ij} and the point X at which g is estimated.

Our approach is first to compute a preliminary approximation, $\tilde{g}$ and then refine it using local linear smoothing within a window. We shall consider a particularly simple preliminary estimator, based on kernel methods, as follows. Write i_n for the integer nearest to nX, let K be a nonnegative, compactly supported, continuously differentiable function, let h_1 equal a constant multiple of $n^{-2/3}$ and put

$$T(j) = \left(nh_1^2\right)^{-1} \sum_k K'\{(j-k)/(nh_1)\}Y(i_n/n, j/n),$$

which is a statistical approximation to the first derivative of $f(i_n/n, \cdot)$ at j/n. Let $\hat{\jmath}$ denote a value which produces a global maximum of $|T|$ in the range $C_1 n \le j \le C_2 n$. Our preliminary estimator of $g(X)$ is $\tilde{g}(X) = \hat{\jmath}/n$.

Next we define an improved estimator. Let $\mathscr{W}$ be a square window of side length $h = h(n)$, with its centre at $(i_n/n, \hat{\jmath}/n)$ and, for the sake of definiteness, its axes aligned with those of the grid. Temporarily make the assumption that within $\mathscr{W}$, f assumes a constant value on either side of a line $\mathscr{L}$. We fit $\mathscr{L}$ by least-squares in the class $\mathscr{M}(C, \mathscr{W})$ of all lines $\mathscr{L}$ that divide $\mathscr{W}$ into two sets of vertices of which the larger has no more than C times the number in the smaller (where $C > 1$ is arbitrary but fixed). Specifically, let $\mathscr{I}_1$ [respectively, $\mathscr{I}_2$] denote the set of vertex coordinates $w = (i/n, j/n)$ in $\mathscr{W}$ that lie above [below] $\mathscr{L}$, let $\Sigma^{(i)}$ denote the sum of $Y(w)$ over all $w \in \mathscr{I}_i$, let

$\bar{Y}_i$ be the corresponding mean and put

$$S(\mathscr{L}) = \sum_{i=1}^{2} \sum{}^{(i)} \{Y(w) - \bar{Y}_i\}^2.$$

Write $\hat{\mathscr{L}}$ for a line that minimizes $S(\mathscr{L})$ among all straight lines in $\mathscr{M}(C,\mathscr{W})$ that do not pass through any vertices. (The minimum is of course not uniquely attained, and any measurable approach to breaking ties allowed.) Write $\hat{g}(X)$ for the ordinate of the point on $\hat{\mathscr{L}}$ with abscissa X. In the theorem we choose h to optimize performance of $\hat{g}$.

THEOREM 3.1. *Assume conditions* (C_f) *and* (C_g), *that X has an absolutely continuous distribution on* $[0,1]$ *and that X is independent of the errors* ε_{ij}. *Let* $0 < \alpha < 4/3$. *If the error distribution satisfies* $E|\varepsilon_{ij}|^t < \infty$ *for some* $t > 4(3-\alpha)/(4-3\alpha)$ *and h is chosen to equal a constant multiple of* $n^{-\alpha/2}$, *then* $\hat{g}(X) - g(X) = O(n^{-\alpha})$ *with probability* 1.

A version of the theorem may be established in the case where the moment generating function of the error distribution is finite in a neighborhood of the origin. Then, $\hat{g}(X) - g(X) = O\{n^{-4/3}(\log n)^{(2/3)+\delta}\}$ for all $\delta > 0$.

4. Proofs.

PROOF OF THEOREM 2.1. We may assume without loss of generality that $\mathscr{L}$ has gradient $u^{-1} \in (0,1)$. (If the gradient exceeds 1, we switch axes, and if it is negative, we reflect in the y axis.) Translate the origin along the x-axis so that the black grid vertex that is furthest to the right on that axis is at $x = 0$. Let $(v,0)$ denote the place where a given line $\mathscr{L}$ intersects the x axis. By choice of origin, $0 \le v \le 1$. For an integer $j \ge 1$ let $(v_j + k_1 + \cdots + k_j, j)$ denote the coordinates of the point at which $\mathscr{L}$ cuts the line given by $y = j$. Here the k_j's are integers and each $0 \le v_j \le 1$. Our assumption that $u > 1$ means that the only values taken by k_j are $\langle u \rangle \ge 1$ and $\langle u \rangle + 1$, where $\langle \cdot \rangle$ denotes the integer part function. Since $\mathscr{L}$ has equation $x = uy + v$, then the k_j's and v_j's satisfy

$$\sum_{i=1}^{j} k_i = \langle uj + v \rangle \quad \text{and} \quad v_j + \sum_{i=1}^{j} k_i = uj + v.$$

Therefore, $v_j = uj + v - \langle uj + v \rangle$, as illustrated in Figure 1.

In the special case where $\mathscr{L}$ passes through the origin, there is a closely related geometric representation [due to Klein (1907)] of the convergents p_n/q_n of u, as follows. Referring to Figure 2, imagine that a thin black thread lies along the line with equation $y = ux$, where $u > 0$. Tie one end of the thread to the origin and the other to a point infinitely remote from the origin. Move to the left that end of the thread at the origin. As it goes, the thread will catch certain vertices above the line. These vertices are those with coordinates $(q_1, p_1), (q_3, p_3), \ldots$. If we move the origin end of the thread to

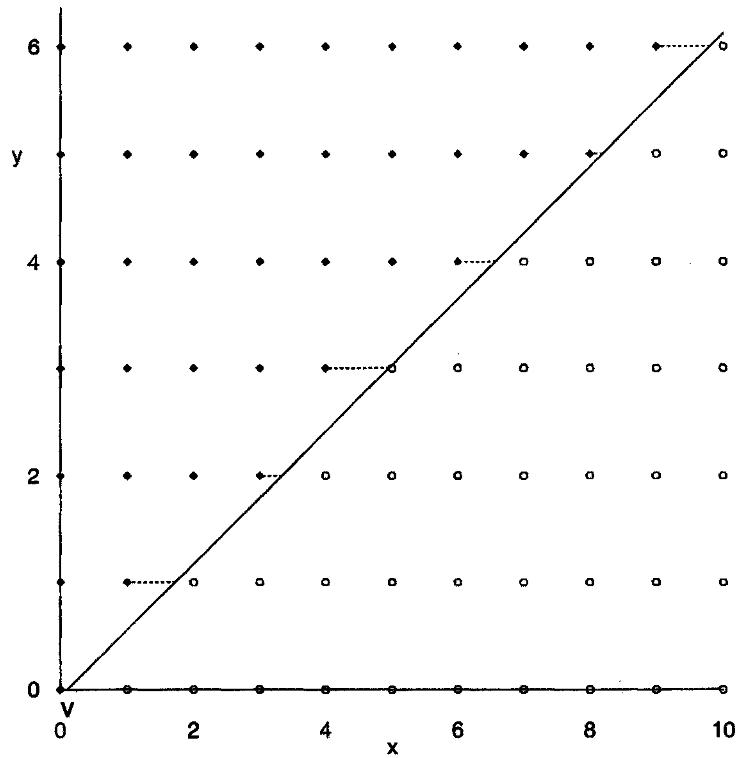

FIG. 1. *Approximation within a section. The line $\mathscr{L}$ is depicted in the case $u = (\sqrt{5} + 1)/2$ and $v = u/16$. The lengths of the horizontal dotted lines, joining vertices to $\mathscr{L}$, are the values of $v_j = uj + v - \langle uj + v \rangle$ for $j = 1, \ldots, 6$.*

the right, it will catch some of the vertices below the line; these have coordinates $(q_0, p_0), (q_2, p_2), \ldots$. Each of the two positions of the thread defines a polygonal path which approximates the line increasingly closely as we move further out.

Let $\mathscr{S}_{\mathscr{W}}$ denote the set of white vertices within $\mathscr{S}$. Then, for any $0 < c_1 < c_2 < \infty$, the horizontal distance from $\mathscr{L}$ of that vertex in $\mathscr{S}_{\mathscr{W}}$ which is closest to $\mathscr{L}$ and has y coordinate j satisfying $c_1 m \le j \le c_2 m$, is

$$d(m, u) = 1 - \max_{c_1 m \le j \le c_2 m} (uj + v - \langle uj + v \rangle). \tag{4.1}$$

[Of course, $d(m, u)$ depends on c_1, c_2 and v as well as m and u, but it is not necessary to stress the former variables.] The validity of (4.1) is illustrated by Figure 2 in the case $c_1 = 1$, $c_2 = 6$, $m = 1$ and $v = 0$. On this occasion, $d(m, u) = p_3 - uq_3$.

Asymptotic properties of the maximum on the right-hand side of (4.1) are determined by properties of rational approximations to the irrational number u. As a prelude to deriving the rate of approximation, we list several

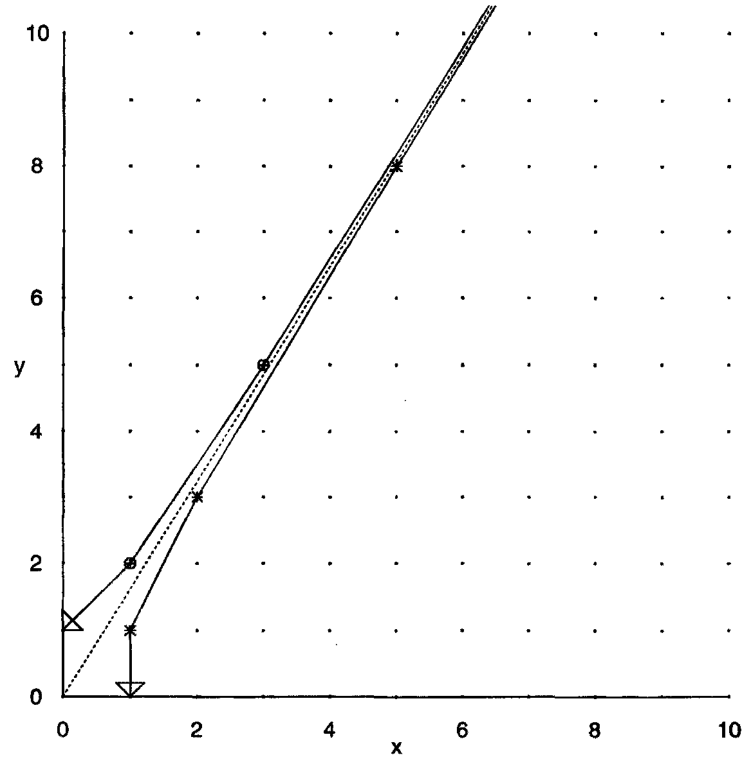

FIG. 2. *Klein diagram. The diagram is drawn in the case $u = (\sqrt{5} + 1)/2$, where the partial denominators $a_n = a_n(u)$ are all equal to 1. The line with this equation is represented by the dotted line, and the positions of the thread (after movement to the left or right) by the unbroken lines. Odd-numbered convergents, above the line, are represented by circles with inscribed crosses, and the even-numbered convergents, below the line, are represented by asterisks.*

properties of convergents:

(4.2) $$\{q_n(q_n + q_{n+1})\}^{-1} < |u - (p_n/q_n)| < (q_n q_{n+1})^{-1},$$

(4.3) $$\inf_{p,\, 1 \le q \le q_n} |u - (p/q)| = |u - (p_n/q_n)|,$$

(4.4) if p and q are relatively prime, and $|u - (p/q)| < (2q^2)^{-1}$, then p/q is a convergent of u.

See, for example, Chapter 9 of Leveque (1956).

Let $0 < c_1 < c_2 < \infty$ be constants, let L be as in Theorem 2.1 and define

$$M_n^{\pm}(u) = n^2 \inf\{\pm[u - (p/q)] : c_1 n \le p \le q \le c_2 n \text{ and } \pm[u - (p/q)] \ge 0\},$$

where the + and − signs are taken, respectively, and

$$M_n(u) = n^2 \inf_{p,\, c_1 n \le q \le c_2 n} |u - (p/q)|.$$

LEMMA 4.1. *Assume $c_2/c_1 > 2$. The following three assertions are equivalent, for any choice of the + and − signs*:

$$\limsup_{n\to\infty} (\log n)^{-1} L(\log n) M_n^{\pm}(u) < \infty \quad \text{for almost all } u, \tag{4.5}$$

$$\liminf_{n\to\infty} (\log n)\{L(\log n)\}^{-1} M_n^{\pm}(u) > 0 \quad \text{for almost all } u, \tag{4.6}$$

$$\sum_{n=1}^{\infty} n^{-1} L(n) < \infty. \tag{4.7}$$

LEMMA 4.2. *Assume $c_2/c_1 > 2$. If u is irrational, then the three following assertions are equivalent, for any choice of the + and − signs*:

$$\limsup_{n\to\infty} M_n^{\pm}(u) < \infty, \qquad \liminf_{n\to\infty} M_n^{\pm}(u) > 0, \quad u \text{ is badly approximable.}$$

We derive only the first lemma in detail, since it is the more difficult. And in both cases, for brevity, we establish the version of the lemma for $M_n^{\pm}$ replaced by M_n. A proof in the case of $M_n^{\pm}$ is similar, making use of the fact that $p_{2n}/q_{2n} \uparrow u$ while $p_{2n+1}/q_{2n+1} \downarrow u$; see Figure 4.2, and also Schmidt (1980), page 11. For the case of $M_n^{\pm}$ we do need more detailed versions of certain results in number theory, in particular Theorem 30 of Khintchine (1963), applying, respectively, to the even and odd subsequences of $\{a_n\}$ (or, virtually equivalently, of $\{q_{n+1}/q_n\}$). However, these may be derived by modifying arguments used to established the results for the full sequence.

PROOF OF LEMMA 4.1. Suppose $|u - (p/q)| < (2q^2)^{-1}$ and $q_n < q < q_{n+1}$. Then either p and q have a common factor, or p/q is a convergent of u [see (4.4)]. In view of the range of values allowed for q the latter cannot be true, and so (a) p and q have a greatest common divisor $k > 1$, and (b) with $p' = p/k$ and $q' = q/k$, p'/q' is a convergent and $q' \le q_n$. Let p''/q'' denote the next convergent; then $q'' \le q_{n+1}$. Hence, by (4.2),

$$|u - (p/q)| = |u - (p'/q')| > (2q'q'')^{-1} \ge (2q_n q_{n+1})^{-1}.$$

Therefore,

$$\inf_{p,\, q_n < q < q_{n+1}:\, |u-(p/q)| < (2q^2)^{-1}} |u - (p/q)| > (2q_n q_{n+1})^{-1}.$$

Using properties (4.2) and (4.3), we may extend this to all $q \le q_n$:

$$\inf_{p,\, 1 \le q < q_{n+1}:\, |u-(p/q)| < (2q^2)^{-1}} |u - (p/q)| > (2q_n q_{n+1})^{-1}. \tag{4.8}$$

If p, q are integers such that $p/q \ne p_n/q_n$ then $|p_n q - p q_n| \ge 1$, or equivalently, $|(p_n/q_n) - (p/q)| \ge 1/(q q_n)$. By property (4.2) of convergents, $|u - (p_n/q_n)| < (q_n q_{n+1})^{-1}$. Therefore, if $p/q \ne p_n/q_n$ and $1 \le q \le q_{n+1}$,

$$\begin{aligned} |u - (p/q)| &\ge |(p_n/q_n) - (p/q)| - |u - (p_n/q_n)| \\ &> (q q_n)^{-1} - (q_n q_{n+1})^{-1} = (q_{n+1} - q)/(q q_n q_{n+1}). \end{aligned}$$

It follows that for any $\varepsilon \in (0,1)$,

$$\inf_{p,\, 1 \le q \le (1-\varepsilon) q_{n+1}} |u - (p/q)| > \varepsilon (1-\varepsilon)^{-1} (q_n q_{n+1})^{-1}. \tag{4.9}$$

Let $\varepsilon_1, \varepsilon_2 \in [0,1)$ with $\varepsilon_1 + \varepsilon_2 < 1$, and let $m > q_n$ be an integer. Provided $m/q_n > (1 - \varepsilon_1 - \varepsilon_2)^{-1}$, the width of interval $(\varepsilon_1 m/q_n, (1-\varepsilon_2)m/q_n)$ is strictly greater than 1, and so the interval contains at least one integer, k say. Necessarily, $\varepsilon_1 m \le k q_n \le (1-\varepsilon_2) m$, and by (4.2), $|u - (kp_n/kq_n)| < (q_n q_{n+1})^{-1}$. Hence,

$$\inf_{p,\, \varepsilon_1 m \le q \le (1-\varepsilon_2) m} |u - (p/q)| < (q_n q_{n+1})^{-1}. \tag{4.10}$$

Additionally, by (4.2), (4.10) holds with $\varepsilon_1 = 1$, $\varepsilon_2 = 0$ and $m = q_n$.

Define $\psi(n) = n^{-1} L(n)$. Recall that $a_n = a_n(u)$ is the nth partial denominator in a continued fraction expansion of u: $u = [a_0; a_1, a_2, \dots]$. If along a subsequence we have either $q_{n+1}/q_n \to \infty$ or $a_{n+1} \to \infty$, then $q_{n+1}/q_n \sim a_{n+1}$ along that subsequence; see for example Khintchine (1963), pages 12, 13. Furthermore, neither q_{n+1}/q_n nor a_{n+1} is less than 1. Therefore, $q_{n+1}/q_n \asymp a_{n+1}$, and so

$$\limsup_{n \to \infty} \psi(n) q_{n+1}/q_n < \infty \tag{4.11}$$

if and only if $\limsup_{n\to\infty} \psi(n) a_{n+1} < \infty$. For almost all u, the latter condition is equivalent to $\Sigma \psi(n) < \infty$, that is, to (4.7); see, for example, Theorem 30 of Khintchine [(1963), page 71], taking Khintchine's $\phi(n)$ to be our $\psi(n)^{-1}$. Furthermore, for almost all choices of u, $\log q_n = \log q_n(u) \sim Cn$, where $C = \pi^2/(12 \log 2)$ [Khintchine (1935), Lévy (1937), page 320]. Therefore, for almost all u, (4.11) is equivalent to

$$\limsup_{n \to \infty} \psi(\log q_n) q_{n+1}/q_n < \infty. \tag{4.12}$$

Hence, it suffices to derive the version of Lemma 4.1 in which (4.7) is replaced by (4.12) (and $M_n^{\pm}$ is replaced by M_n).

First we prove the following result, which we call (A): for all irrational u, condition (4.12) is implied by each of

$$\limsup_{n \to \infty} \psi(\log n) M_n < \infty, \tag{4.13}$$

$$\liminf_{n \to \infty} \psi(\log n)^{-1} M_n > 0, \tag{4.14}$$

Now, condition (4.12) fails if and only if there exists a subsequence $\{n_k\}$ diverging to infinity, such that

$$\psi(\log q_{n_k}) q_{n_k+1}/q_{n_k} \to \infty. \tag{4.15}$$

Taking m_k equal to the integer part of $\delta q_{n_k+1}/c_2$, where $0 < \delta < 1$, and applying (4.9) with $1 - \varepsilon \in (\delta, 1)$, we see that (4.15) implies $\psi(\log m_k) M_{m_k} \to \infty$. Therefore, (4.13) fails. Alternatively, taking m_k equal to the smallest integer exceeding $c_2^{-1}q_{n_k}$, and applying the right-hand inequality at (4.2), we deduce from (4.15) first that $(q_{n_k}/q_{n_{k+1}})\psi(\log m_k)^{-1} \to 0$ and then that $\psi(\log m_k)^{-1}M_{m_k} \to 0$. Therefore, (4.14) fails. This proves result (A).

Next we establish a converse, which we call result (B): for all irrational u, the failure of either (4.13) or (4.14) implies that of (4.12). If (4.13) fails then we may choose $m_k \to \infty$ such that

$$\psi(\log m_k)M_{m_k} \to \infty. \tag{4.16}$$

Let n_k be the largest integer such that $q_{n_k} \le c_2 m_k$. Then by (4.3), and the right-hand inequality in (4.2), $M_{m_k} = O(m_k^2/q_{n_k}q_{n_k+1})$. But $q_{n_k+1} > c_2 m_k$, and so $M_{m_k} = O(q_{n_k+1}/q_{n_k})$. Therefore, $\psi(\log m_k)M_{m_k} \to \infty$ implies that $q_{n_k+1}/q_{n_k} \to \infty$. Let $\varepsilon \in (0,1)$ be so small that $c_1 < (1-\varepsilon)c_2$, and let l_k equal the integer part of $(1-\varepsilon)c_2^{-1}q_{n_k+1}$. If $m_k < l_k$ then (4.16) holds with l_k replacing m_k. [To appreciate why, note that if $m_k < l_k$ then (1) $\psi(\log m_k)m_k^2 < 2\psi(\log l_k)l_k^2$ for all sufficiently large k; (2) by (4.10) with $m = \text{const.}\, m_k$, using (in the case $q_{n_k} < c_1 m_k$) the fact that $c_2/c_1 > 2$, we see that $m_k^{-2}M_{m_k} \le (q_{n_k}q_{n_k+1})^{-1}$; and (3) by (4.9), $(q_{n_k}q_{n_k+1})^{-1}$ is dominated by a constant multiple of $l_k^{-2}M_{l_k}$. It follows that $\psi(\log m_k)M_{m_k}$ is dominated by a constant multiple of $\psi(\log l_k)M_{l_k}$, and so (4.16) implies the same result with l_k replacing m_k.] Therefore, we may assume without loss of generality that $l_k \le m_k$. In this case we have

$$M_{m_k} \le (m_k/l_k)^2 M'_{l_k} < \left[\left(c_2^{-1}q_{n_k+1}\right)/\left\{(1-\varepsilon)c_2^{-1}q_{n_k+1} - 1\right\}\right]^2 M'_{l_k} = O(M'_{l_k}),$$

where

$$M'_{l_k} = l_k^2 \inf_{p,\, c_1 m_k \le q \le c_2 l_k} |u - (p/q)|.$$

The ratio of upper and lower end points of the interval over which the latter infimum is taken, is

$$c_2 l_k/(c_1 m_k) \sim (1-\varepsilon)q_{n_k+1}/(c_1 m_k)$$
$$> (1-\varepsilon)q_{n_k+1}/\left(c_1 c_2^{-1} q_{n_k+1}\right) = (1-\varepsilon)c_2/c_1 > 1.$$

Thus, in view of (4.10) with $m = \text{const.}\, q_{n_k+1}$,

$$M'_{l_k} = O\left(l_k^2/q_{n_k}q_{n_k+1}\right) = O\left(q_{n_k+1}/q_{n_k}\right).$$

Furthermore, since $q_{n_k} \le c_2 m_k$ then $\psi(\log m_k) \le 2\psi(\log q_{n_k})$ for large k. Therefore,

$$\psi(\log m_k)M_{m_k} = O\left\{\psi(\log q_{n_k})M'_{l_k}\right\} = O\left\{\psi(\log q_{n_k})q_{n_k+1}/q_{n_k}\right\}.$$

Hence, (4.16) implies that

$$\psi(\log q_{n_k})q_{n_k+1}/q_{n_k} \to \infty,$$

and so (4.12) fails.

We complete the derivation of result (B) by showing that if (4.14) fails, then so too does (4.12). When (4.14) fails, there exists a subsequence $m_k \to \infty$

such that

$$\psi(\log m_k)^{-1} M_{m_k} \to 0. \tag{4.17}$$

In particular, $M_{m_k} \to 0$ and so

$$m_k^2 \inf_{p,\, 1 \le q \le m_k} |u - (p/q)| \to 0. \tag{4.18}$$

Let n_k be the largest integer such that $q_{n_k} \le c_2 m_k$. Result (4.17) implies that

$$\psi(\log m_k)^{-1} m_k^2 \inf_{p,\, 1 \le q < q_{n_k+1}} |u - (p/q)| \to 0.$$

Thus, (4.8) and (4.18) give

$$\psi(\log m_k)^{-1} m_k^2 (q_{n_k} q_{n_k+1})^{-1} \to 0.$$

But $m_k \ge c_2^{-1} q_{n_k}$ and so $\psi(\log m_k)^{-1} \ge \frac{1}{2}\psi(\log q_{n_k})^{-1}$ for large k. Therefore, $\psi(\log q_{n_k})^{-1} q_{n_k}/q_{n_k+1} \to 0$, implying that (4.12) fails and establishing result (B).

Results (A) and (B) imply that for almost all u, (4.12), (4.13) and (4.14) are equivalent. The latter two conditions are the versions of (4.5) and (4.6) when M_n replaces $M_n^{\pm}$. We have already shown that (4.12) is equivalent to (4.7) and so we have established the version of Lemma 4.1 with the aforementioned interchange of M_n and $M_n^{\pm}$. □

Proof of Lemma 4.2. Arguing as in the proof of Lemma 4.1, we may show that each of the first two assertions in Lemma 4.2 is equivalent to $\sup_n q_{n+1}(u)/q_n(u) < \infty$. This in turn is equivalent to $\sup_n a_n(u) < \infty$ [see, e.g., Khintchine (1963), pages 12, 13], and so to u being BA.

Lemmas 4.1 and 4.2 are employed to derive parts (a) and (b), respectively, of Theorem 2.1. We illustrate the argument by considering the case of part (b). The first step in proving (2.1) is to derive the following result. Let $\mathscr{S}^{(0)}$ denote the set of vertices that have their y coordinates in $[c_1 m, c_2 m]$. If $v = 0$ (i.e., if the line $\mathscr{L}$ passes through a black vertex), and if c_2/c_1 is sufficiently large (the size depending only on u), then the white vertex in $\mathscr{S}^{(0)}$ nearest to $\mathscr{L}$ is distant $O(m^{-1})$ from $\mathscr{L}$. We call this result (R_{white}). Similarly, the version (R_{black}) for black vertices is true when $v = 1$ (i.e., when the line passes through a white vertex).

We derive only (R_{white}), which is equivalent to showing that

$$d(m, u) = O(m^{-1}) \tag{4.19}$$

in the special case where $v = 0$. Let $\mathscr{E}_1(c, u)$ denote the event that $d(m, u) \le cm^{-1}$ for all but a finite number of values of m. Given $c_3 > 0$, $c_4 > 1$ and an integer $b > 1$, define

$$\mathscr{I}_j(b, c_3, c_4) = \bigcup_{j \le j_1 \le c_4 j} \bigcup_{i=j_1}^{bj_1} \left(\frac{i}{j_1} - \frac{c_3}{j_1^2}, \frac{i}{j_1} \right).$$

For $1 < u < b$, let $\mathscr{E}_2(b, c_3, c_4, u)$ be the event that $u \in \mathscr{I}_j(b, c_3, c_4)$ for all but a finite number of values of j. Then, provided $c_3/c_1 < c$ and $c_4 < c_2/c_1$, $\mathscr{E}_2(b, c_3, c_4, u) \subseteq \mathscr{E}_1(c, u)$. In view of Lemma 4.1, noting particularly the one-sided character of M_n^+ and M_n^-, if u is BA and $c_4 > 0$ is sufficiently large, then $u \in \mathscr{I}_j(b, 1, c_4)$ for all sufficiently large j. Therefore, if c_2/c_1 is sufficiently large and $c > 1/c_1$, $d(m, u) \leq cm^{-1}$ for all sufficiently large m. This proves (R_{white}).

The next step in deriving (2.1) is to prove that (4.19) holds for general v. Put $c_5 = (c_2 + c_1)/2$ and let $\mathscr{S}^{(1)}$ and $\mathscr{S}^{(2)}$ denote the sets of vertices that have their y coordinates in $[c_1 m, c_5 m]$ and $(c_5 m, c_2 m]$, respectively. Let V be the black vertex in $\mathscr{S}^{(0)}$ that is nearest to $\mathscr{L}$. (In this part of the proof we shall measure all distances horizontally.) Without loss of generality, $V \in \mathscr{S}^{(1)}$. Let $\mathscr{L}'$ denote the line parallel to $\mathscr{L}$ and passing through V, let V' be the vertex in $\mathscr{S}^{(2)}$ that is below $\mathscr{L}'$ and nearest to $\mathscr{L}'$ of all vertices with this property and let $\mathscr{L}''$ denote the line passing through V' and parallel to $\mathscr{L}$. By definition of V, the vertex V' must be white, since otherwise it would be a black vertex in $\mathscr{S}^{(0)}$ closer to $\mathscr{L}$ than V. By construction of $\mathscr{L}'$ and $\mathscr{L}''$, $\mathscr{L}$ lies between these lines. And by (R_{white}), if c_2/c_1 is sufficiently large (depending only on u, not on v), V' must be within $O(m^{-1})$ of $\mathscr{L}'$. It follows from the last of these three properties that $\mathscr{L}'$ and $\mathscr{L}''$ are $O(m^{-1})$ apart and thence from the second property that V' is within $O(m^{-1})$ of $\mathscr{L}$. Therefore, there exists a white vertex in $\mathscr{S}^{(0)}$ that lies within $O(m^{-1})$ of $\mathscr{L}$. Similar arguments involving R_{black} show that there exists a black vertex within the same distance of $\mathscr{S}^{(0)}$.

Summarizing, if c_2/c_1 is sufficiently large, then the black and white vertices in $\mathscr{S}^{(0)}$ that are nearest to $\mathscr{L}$ are distant $O(m^{-1})$ from $\mathscr{L}$. Applying this result to two small squares of size $pm \times pm$ (with $0 < p < 1$) positioned where $\mathscr{L}$ enters $\mathscr{S}$ on the left or below, and leaves $\mathscr{S}$ on the right or above, respectively, we see that the greatest horizontal distance between those parts of $\mathscr{L}$ and $\hat{\mathscr{L}}$ that lie within $\mathscr{S}$ is of order m^{-1}, and similarly for vertical distances. This establishes (2.1).

We only outline the derivation of (2.2), since it is similar to that of (2.1). In place of (R_{white}) we may prove that if $v = 0$ then the white vertex in $\mathscr{S}^{(0)}$ nearest to $\mathscr{L}$ is distant at least $O(m^{-1})$ from $\mathscr{L}$. (Again, distance is measured horizontally.) Using Lemma 4.1, for each BA irrational number u, one can show that $d(m, u) \geq \text{const.} \times m^{-1}$. We call this result (R'_{white}) and its analogue (R'_{black}). Next, define V, V', $\mathscr{L}'$ and $\mathscr{L}''$ as in the proof of (2.1). By (R'_{white}), the parallel lines $\mathscr{L}'$ and $\mathscr{L}''$ are at least a constant multiple of m^{-1} apart by their definition, there are no vertices in $\mathscr{S}^{(0)}$ that lie between them and, as before, $\mathscr{L}$ lies between them. Hence, there exists a rectangular prism with its long sides parallel to $\mathscr{L}$ and its short sides horizontal and of length at least a constant multiple of m^{-1}, such that any line $\mathscr{L}'''$ that passes through the prism produces exactly the same vertex color pattern in $\mathscr{S}^{(0)}$ as did $\mathscr{L}$. This proves (2.2).

By working with subsequences, we may prove that both (2.1) and (2.2) fail if u is not BA. □

PROOF OF THEOREM 3.1. We condition on $X = x$, where x is chosen so that a straight line $\mathscr{L}$ with gradient $u = g'(x)$ may, in the context of Theorem 2.1, be approximated with error $O\{m^{-1}(\log m)^{1+\varepsilon}\}$ for all $\varepsilon > 0$. Since the distribution of X is continuous, then with probability 1, the value of X has this property.

STEP 1 [Performance of $\tilde{g}(x)$]. The estimator $\tilde{g}$ need only converge to g a little faster than the square root of the rate of convergence of $\hat{g}$: $\tilde{g}(x) - g(x) = o(n^{-\alpha/2})$,

$$\tilde{g}(x) - g(x) = O(n^{-2/3}) \tag{4.20}$$

as $n \to \infty$, with probability 1.

We give this proof only in outline. Using the Taylor expansion and integral approximation to series and observing that h_1 is of size $n^{-2/3}$, it may be shown that $\tau(j) \equiv E\{T(j)|g\} = \tau_1(j) + r(j)$, where

$$\tau_1(j) = h_1^{-1} f_2\{i_n/n, g(i_n/n)\} K\big[\{(j/n) - g(i_n/n)\}/h_1\big],$$
$$r(j) = O\Big\{\big(nh_1^2\big)^{-1} + \big(n^2h_1^3\big)^{-1} + 1\Big\} = O(n^{1/3}).$$

We note that, in case $\alpha > 0$, $E|\varepsilon_{ij}|^t < \infty$ for some $t > 3$ and we employ a strong approximation to partial sums of independent and identically distributed random variables [available from Shorack and Wellner (1986), pages 60, 61]. Using Euler's method of summation (the summation analogue of integration by parts), we may prove that there exists a Gaussian process ξ, defined on the integers, with zero mean and the same covariance structure as $T(j)$ conditional on g, such that for some $\eta > 0$,

$$\sup_j |T(j) - \tau(j) - \xi(j)| = O\big(n^{-(2/3)-\delta} h_1^{-2}\big) = O\big(n^{(2/3)-\eta}\big),$$

with probability 1. Fernique's lemma [Marcus (1970)] may be applied to show that

$$\sup_j |\xi(j)| = O\Big[\Big\{\big(nh_1^3\big)^{-1} \log n\Big\}^{1/2}\Big] = O\big\{(n \log n)^{1/2}\big\}$$

with probability 1. Combining the results so far in this paragraph, we see that for some $\eta > 0$,

$$T(j) = \tau_1(j) + O\big(n^{(2/3)-\eta}\big).$$

Therefore, since K vanishes outside a compact interval, the value $\hat{\jmath}$ that maximizes $|T(j)|$ satisfies $g(i_n/n) - (\hat{\jmath}/n) = O(h_1)$. However, by definition of i_n and since g is differentiable, $g(i_n/n)$ is within $O(n^{-1})$ of $g(x)$. Hence, $g(x) - (\hat{\jmath}/n) = O(h_1)$, which is equivalent to (4.20).

STEP 2 (Completion). Note that the window width h used to construct $\tilde{g}$ is taken to be the square root of the claimed convergence rate of $\hat{g}$ to g and that, by Step 1, $\tilde{g}$ converges to g at rate $o(h)$. Therefore, the point $(i_n/n, g(x))$

lies asymptotically at the center of $\mathscr{W}$, in the sense that its distance from the true center to $(i_n/n, g(x))$, divided by the width of $\mathscr{W}$, converges to zero with probability 1 as $n \to \infty$.

Let $\mathscr{L}_1$ denote a line across $\mathscr{W}$ which, in the absence of noise, would minimize the sum of squares $S(\mathscr{L})$. (Then $\mathscr{L}_1$ is constrained to cross $\mathscr{W}$ within a certain region; it is not uniquely defined.) Let $\mathscr{L}_2$ be any other line across $\mathscr{W}$. We shall address the case where $\mathscr{L}_1$ and $\mathscr{L}_2$ intersect inside $\mathscr{W}$; the case where they do not intersect there is a little simpler.

The lines $\mathscr{L}_1$ and $\mathscr{L}_2$ divide $\mathscr{W}$ into four regions, $\mathscr{R}_{11}$, $\mathscr{R}_{12}$, $\mathscr{R}_{21}$ and $\mathscr{R}_{22}$, with $\mathscr{R}_{11}$ opposite $\mathscr{R}_{12}$ and $\mathscr{R}_{21}$ opposite $\mathscr{R}_{22}$, and $\mathscr{L}_i$ divides $\mathscr{W}$ into two regions $\mathscr{S}_{i1}$ and $\mathscr{S}_{i2}$. Select notation so that (1) the area of $\mathscr{R}_{21} \cup \mathscr{R}_{22}$ is less than or equal to that of $\mathscr{R}_{11} \cup \mathscr{R}_{12}$ and (2) $\mathscr{S}_{1i} = \mathscr{R}_{1i} \cup \mathscr{R}_{2i}$ and $\mathscr{S}_{2i} = \mathscr{R}_{1i} \cup \mathscr{R}_{2,3-i}$. Let N_{ij} denote the number of vertices in $\mathscr{S}_{ij}$, let T_i (respectively, D_i) be the sum of $Y(w)$ over all vectors w of vertices in $\mathscr{R}_{1i}$ ($\mathscr{R}_{2i}$) and put $\overline{Y}_{ij} = N_{ij}^{-1}(T_j + D_i)$. Then

$$\begin{aligned}
S(\mathscr{L}_2) &- S(\mathscr{L}_1) \\
&= N_{11}\overline{Y}_{11}^2 + N_{12}\overline{Y}_{12}^2 - \left(N_{21}\overline{Y}_{21}^2 + N_{22}\overline{Y}_{22}^2\right) \\
&= (N_{21} - N_{11})(N_{11}N_{21})^{-1}T_1^2 + (N_{22} - N_{12})(N_{12}N_{22})^{-1}T_2^2 \\
&\quad + (N_{12} + N_{11})(N_{12}N_{11})^{-1}D_1^2 - (N_{22} + N_{21})(N_{21}N_{22})^{-1}D_2^2 \\
&\quad + 2T_1\left(N_{11}^{-1}D_1 - N_{21}^{-1}D_2\right) + 2T_2\left(N_{12}^{-1}D_1 - N_{22}^{-1}D_2\right).
\end{aligned} \tag{4.21}$$

Write $T_i = t_i + \tau_i$ and $D_i = d_i + \delta_i$, where τ_i and δ_i denote the respective sums of errors ε_{ij}, and t_i and d_i are the respective sums of terms involving f. Then

$$S(\mathscr{L}_2) - S(\mathscr{L}_1) = s(\mathscr{L}_1, \mathscr{L}_2) + \sigma(\mathscr{L}_1, \mathscr{L}_2),$$

where $s(\mathscr{L}_1, \mathscr{L}_2)$ has the same formula as $S(\mathscr{L}_2) - S(\mathscr{L}_1)$ except that (T_i, D_i) is replaced by (t_i, d_i) and $\sigma(\mathscr{L}_1, \mathscr{L}_2)$ is defined as the difference.

Let $\nu = \nu(\mathscr{L}_1, \mathscr{L}_2)$ denote the number of vertices in $\mathscr{R}_{21} \cup \mathscr{R}_{22}$. Then for all sufficiently large n, $s(\mathscr{L}_1, \mathscr{L}_2) \geq B_1\nu$, where $B_1, B_2, \dots$ are positive constants depending on f and by (4.21),

$$\begin{aligned}
&|\sigma(\mathscr{L}_1, \mathscr{L}_2)| \\
&\quad \leq B_2 \sum_{i=1}^{2} \left[\nu\left\{N^{-2}\left(\tau_i^2 + |t_i\tau_i|\right) + N^{-1}|\tau_i|\right\} + N^{-1}\left(\delta_i^2 + |d_i\delta_i|\right) + |\delta_i|\right],
\end{aligned}$$

where N equals the total number of vertices in $\mathscr{W}$. (The assumptions in the theorem and the result of Step 1 imply that for all sufficiently large n and for each pair i, j, N_{ij} does not exceed a constant multiple of N.) Therefore, since $|d_i| \leq B_3 N$ and $\nu \leq N$,

$$\begin{aligned}
p(\nu_0) &\equiv P\{S(\mathscr{L}_2) - S(\mathscr{L}_1) \leq 0 \text{ for some } \mathscr{L}_2 \text{ with } \nu(\mathscr{L}_1, \mathscr{L}_2) \geq \nu_0\} \\
&\leq \sum_{(\nu_0)} \sum_{i=1}^{2} \{P(|\tau_i| > B_4 N) + P(|\delta_i| > B_4\nu)\},
\end{aligned} \tag{4.22}$$

where $\Sigma_{(\nu_0)}$ and $\sup_{(\nu_0)}$ denote, respectively, summation and supremum over all the different partitions of vertices in $\mathscr{W}$ that arise from different choices of $\mathscr{L}_2$ such that $\nu(\mathscr{L}_1, \mathscr{L}_2) \geq \nu_0$. There are at most N^2 summands addressed by $\Sigma_{(\nu_0)}$, and so

$$p(\nu_0) \leq 2N^2 \sum_{i=1}^{2} \sup_{(\nu_0)} \{P(|\tau_i| > B_4 N) + P(|\delta_i| > B_4 \nu)\}.$$

If $E|\varepsilon_{ij}|^t < \infty$, then the right-hand side equals $O(N^2 \nu_0^{-t/2})$:

$$p(\nu_0) = O(N^2 \nu_0^{-t/2}). \tag{4.23}$$

Let h equal a constant multiple of $n^{-\alpha/2}$, choose

$$t > t_0 = 4(3 - \alpha)/(4 - 3\alpha)$$

and define $u = 3 - (1 - \frac{1}{2}\alpha)^{-1}$ and ν_0 to equal the integer part of $(nh)^u$. Then, by (4.23) and since $N = O\{(nh)^2\}$, we have, for some $\delta > 0$,

$$p(\nu_0) = O\{(nh)^{4-(tu)/2}\} = O\{(nh)^{4-(t_0 u)/2} n^{-\delta}\} = O(n^{-1-\delta}).$$

Since this quantity is summable in n, we have by (4.22) and the Borel–Cantelli lemma that

$$P\{S(\mathscr{L}_2) - S(\mathscr{L}_1) > 0 \text{ for all } \mathscr{L}_2 \text{ with } \nu(\mathscr{L}_1, \mathscr{L}_2) \geq \nu_0, \text{ and all sufficiently large } n\} = 1. \tag{4.24}$$

Therefore, if $\hat{\mathscr{L}}$ minimizes $S(\cdot)$, then with probability 1 $\nu(\mathscr{L}_1, \hat{\mathscr{L}}) \leq \nu_0$ for all sufficiently large n. Now, $D(\mathscr{L}_1, \hat{\mathscr{L}})hn^2 = O\{\nu(\mathscr{L}_1, \hat{\mathscr{L}})\}$, and so

$$D(\mathscr{L}_1, \hat{\mathscr{L}}) = O\{(n^2 h)^{-1} \nu(\mathscr{L}_1, \hat{\mathscr{L}})\} = O\{(n^2 h)^{-1} \nu_0\} = O(n^{-\alpha}),$$

with probability 1. Hence, $\hat{g}(x)$ (the ordinate of the point on $\hat{\mathscr{L}}$ with abscissa x) is $O(n^{-\alpha})$ from the point on $\mathscr{L}_1$ with abscissa x. A simpler, geometric argument shows that the latter point is $O(h^2) = O(n^{-\alpha})$ from $g(x)$, and so $\hat{g}(x) - g(x) = O(n^{-\alpha})$ with probability 1, as had to be proved. □

Acknowledgments. The authors are grateful to Drs. M. Berman, M. J. Kocan and M. F. Newman for helpful discussions.

REFERENCES

BRADY, M. and ASADA, H. (1984). Smoothed local symmetrics and their implementation. *Internat. J. Robotics Res.* **3** 36–60.

DORST, L. and SMEULDERS, A. W. M. (1984). Discrete representation of straight lines. *IEEE Trans. Pattern Anal. Machine Intell.* **6** 450–463.

DORST, L. and SMEULDERS, A. W. M. (1986). Best linear unbiased estimators for properties of digitized straight lines. *IEEE Trans. Pattern Anal. Machine Intell.* **8** 276–282.

DORST, L. and SMEULDERS, A. W. M. (1987). Length estimators for digitized contours. *Comput. Vision Graphics Image Process.* **40** 311–333.

FREEMAN, H. (1970). Boundary encoding and processing. In *Picture Processing and Psychopictorics* (B. S. Lipkin and A. Rosenfeld, eds.) 241. Academic Press, New York.

GROEN, F. C. A. and VERBEEK, P. W. (1978). Freeman-code probabilities of object boundary quantized contours. *Comput. Vision Graphics Image Process.* **7** 391–402.

HASTIE, T. and LOADER, C. (1993). Local regression: automatic kernel carpentry. *Statist. Sci.* **8** 120–143.

HUNG, S. Y. H. (1985). On the straightness of digital arcs. *IEEE Trans. Pattern Anal. Machine Intell.* **2** 203–215.

KHINTCHINE, A. YA. (1935). Zur metrischen Kettenbruch-Theorie. *Compositio Math.* **3** 275–285.

KHINTCHINE, A. YA. (1963). *Continued Fractions.* Noordhoff, Groningen.

KLEIN, F. (1907). *Ausgewählte Kapital der Zahlentheorie.* Teubner, Leipzig.

KOPLOWITZ, J. and BRUCKSTEIN, A. M. (1989). Design of perimeter estimators for planar shapes. *IEEE Trans. Pattern Anal. Machine Intell.* **11** 611–622.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statist.* **82**. Springer, Berlin.

KULPA, Z. (1977). Area and perimeter measurement of blobs in discrete binary pictures. *Comput. Vision Graphics Image Process.* **6** 434–454.

LANDAU, U. M. (1987). Estimation of a circular arc center and its radius. *Comput. Vision Graphics Image Process.* **38** 317–326.

LEVEQUE, W. J. (1956). *Topics in Number Theory* **1**. Addison-Wesley, Reading, MA.

LÉVY, P. (1937). *Théorie de l'Addition des Variables Aléatories.* Gauthier-Villars, Paris.

MARCUS, M. B. (1970). A bound for the distribution of the maximum of continuous Gaussian processes. *Ann. Math. Statist.* **41** 305–309.

MCILROY, M. D. (1985). A note on discrete representation of lines. *AT&T Technical J.* **64** 481–490.

PROFITT, D. and ROSEN, D. (1979). Metrication errors and coding efficiency of chain-encoding schemes for the representation of lines and edges. *Comput. Vision Graphics Image Process.* **10** 318–332.

SCHMIDT, W. M. (1980). Diophantine approximation. *Lecture Notes in Math.* **785**. Springer, Berlin.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

WORRING, M. and SMEULDERS, W. M. (1995). Digitized circular arcs: characterization and parameter estimation. *IEEE Trans. Pattern Anal. Machine Intell.* **17** 587–598.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, A.C.T. 0200
AUSTRALIA