The Annals of Statistics
1993, Vol. 21, No. 4, 1948–1966

THE EXPECTED NUMBER OF LOCAL MAXIMA OF A RANDOM FIELD AND THE VOLUME OF TUBES

BY DAVID SIEGMUND[1] AND HEPING ZHANG

Stanford University and Yale University

Using an expression for the expected number of local maxima of a random field, we derive an upper bound for the volume of a tube about a manifold in the unit sphere and show that under certain conditions our bound agrees with the evaluation of the tube volume in Weyl's formula. Applications to tests and confidence regions in nonlinear regression are discussed.

1. Introduction. Hotelling (1939) posed the statistical problem of testing for the presence of a nonlinear term in a linear model. In the case that the nonlinear term contained a single nonlinear parameter, he showed that the significance level of the likelihood ratio test involves the volume of a tube about a curve (actually two tubes about two curves) imbedded in the unit sphere in m-dimensional Euclidean space, where m is the sample size minus the number of linear terms in the model. He then gave a geometric evaluation of the volume of the tube under the assumption that the tube radius, hence the significance level, is small. Hotelling's result states that if the curve is closed the tube volume is essentially the cross-sectional area of the tube multiplied by the arc length of the curve. In a companion paper Weyl (1939) treated the case of an arbitrary number of nonlinear parameters, where the geometric problem involves the volume of a tube about a manifold of dimension equal to the number of nonlinear parameters, say q. An important difference between the one-dimensional and multidimensional results is that Hotelling's formula does not explicitly involve the curvature of the manifold (curve), whereas Weyl's does. Weyl's paper is regarded as classical in differential geometry [cf. Gray (1982)], but until recently neither Hotelling's nor Weyl's paper has had much impact in statistics. See Naiman (1986, 1987, 1990), Johansen and Johnstone (1990), Johnstone and Siegmund (1989), Knowles (1987), Knowles and Siegmund (1989), Sun (1989) and Knowles, Siegmund and Zhang (1991) for recent related statistical research. A completely different application is discussed by Smale (1981), who apparently was under the erroneous impression that in the case of arbitrary radii the Hotelling–Weyl formulas give upper bounds for the tube volume.

Received January 1992; revised November 1992.
[1]Supported in part by National Security Agency Grant MDA 904-89-H-2040 and NSF Grant DMS-91-04432.
AMS 1991 *subject classifications*. Primary 62J02; secondary 53A07.
Key words and phrases. Tube volume, nonlinear regression.

Naiman (1986) noted that the same geometric problem arises in finding simultaneous confidence bands in curvilinear regression, and showed that with the proper interpretation Hotelling's formula, which gives the exact volume for tubes of sufficiently small radii, gives an upper bound for all radii and hence a conservative approximation to the desired probability. A simple geometric description of the situation in one dimension is the following. For a curve which is not closed, the expression given by Hotelling must be trivially modified by adding to it the volume of a sphere, to account for the two hemispherical caps at the ends of the curve. Naiman's result is that if one always adds to Hotelling's evaluation this spherical volume, which is obviously necessary when the curve is not closed, the result provides an upper bound for the volume of the tube for any curve and any tube radius. See also Estermann (1926), where Naiman's result was anticipated from an entirely different viewpoint for tubes about curves in Euclidean space. Johnstone and Siegmund (1989) gave two unified derivations of the Hotelling–Naiman results. One of these involved a probabilistic argument using the concept of upcrossings borrowed from the theory of Gaussian processes [cf. Leadbetter, Lindgren and Rootgen (1983), Chapter 7].

In this paper we give a similar analysis of the volume of a tube about a q-dimensional manifold. In principle, our method is a direct application of a formula for the expected number of local maxima of a random field. However, the situation is much more subtle than in one dimension. Although our expression agrees with Weyl's under certain conditions, in particular whenever Weyl's result is exact, in more than one dimension there does not appear to be any simple, geometrically based modification of Weyl's formula which will yield our upper bound. We give simple examples where Weyl's formula grossly underestimates the tube volume and where it exceeds our upper bound. In one dimension our bound can be better and can be worse than Naiman's.

In Section 2 we derive our upper bound and show that under certain conditions it agrees with Weyl's formula. Section 3 contains two simple examples which indicate some of the geometric subtlety of the multidimensional case. In Section 4 we discuss numerical evaluation of our bound, and numerical examples involving some two-dimensional manifolds are given in Section 5.

Section 6 is concerned with the related problem of confidence regions in nonlinear regression along the lines of Knowles, Siegmund and Zhang (1991), who discussed models involving only one nonlinear parameter.

REMARK. We are using the expression "Weyl's formula" to refer to displays (10) and (16) of Weyl (1939). Weyl himself (with understatement) regarded these as "hardly ... more than what could have been accomplished by any student in a course of calculus," and went on to show that (16) could be expressed intrinsically, that is, entirely in terms of the metric tensor of the manifold. Our result is not intrinsic, and our methods make no contribution to understanding this last, much deeper result, which from a geometric point of view is more properly deserving the designation "Weyl's formula."

2. An upper bound for the tube volume. Let S^{m-1} denote the unit sphere in m-dimensional Euclidean space. For $1 \le q < m - 1$ let M be a smooth q-dimensional manifold imbedded in S^{m-1}. We suppose that M is given locally as the image of a smooth function $\gamma = \gamma(t)$ defined on a suitable subset T of q-dimensional Euclidean space. For most statistical applications the manifold is defined globally by γ, and we shall not concern ourselves with technicalities associated with manifolds requiring more than one coordinate patch for their definition. The tube about γ of geodesic radius ϕ, $0 < \phi \le \pi$, is the set of all points $u \in S^{m-1}$ such that $\langle \gamma, u \rangle > \cos(\phi)$ for some $\gamma \in M$. We shall be interested in the volume of the tube of geodesic radius $\phi = \cos^{-1}(w)$ about γ, or equivalently

$$P\Big\{ \sup_{\gamma \in M} \langle \gamma, U \rangle > w \Big\}, \tag{1}$$

where U is uniformly distributed on S^{m-1}.

Since boundary points of the manifold require special treatment, we assume initially that M has no boundary, or what is equivalent for our purposes that the event in (1) is intersected with those points $u \in S^{m-1}$ for which the closest point on the manifold M is an interior point. Let $Z(t) = \langle \gamma(t), U \rangle$, $t \in T$, and let N_w denote the number of local maxima of $Z(t)$ at which $Z(t) > w$. Then (1) is equal to $P\{N_w \ge 1\}$. Note that

$$P\{N_w \ge 1\} \le E(N_w).$$

Our upper bound for (1) is provided by the following evaluation of $E(N_w)$.

PROPOSITION 1. *For a manifold M without boundary,*

$$E(N_w) = \int_T E\big\{ | -\ddot{Z} |;\, Z > w,\, \dot{Z} = 0,\, \ddot{Z} < 0 \big\}\, dt, \tag{2}$$

where $\dot{Z}$ and $\ddot{Z}$ are respectively the gradient and Hessian of Z, $| -\ddot{Z}|$ is the determinant of $-\ddot{Z}$ and the notation $\ddot{Z} < 0$ means that $\ddot{Z}$ is negative definite.

Proposition 1 is well known when Z is a well-behaved smooth Gaussian process. See, for example, Adler (1981), pages 123–124, or Belyayev (1972). For the process $Z(t) = \langle \gamma(t), U \rangle$ of interest here, there are some additional technicalities, which are already apparent in Johnstone and Siegmund's (1989) derivation of the expected number of upcrossings of the level w in the comparatively simple case that $q = 1$. Since these details would lead to a lengthy technical digression and add no important insight to what follows, we shall not discuss them further.

Under the assumption that the tube radius is sufficiently small that no overlap occurs, Weyl (1939) gave several expressions for the volume of the tube about M, which obviously equals the product of the probability (1) and

$$\text{volume}(S^{m-1}) = 2\pi^{m/2}/\Gamma(m/2). \tag{3}$$

We now show that under assumption (11) below which includes all cases where Weyl's formula is exact, Weyl's formula equals the product of (3) and the right-hand side of (2).

At $\gamma(t) \in M$ consider the basis for $\mathbb{R}^m$ given by $\dot{\gamma}_i$, the partial derivative of γ with respect to t_i, $i = 1, \ldots, q$, and $n(\nu)$, $\nu = 1, \ldots, m-q$, where the $n(\nu)$ are mutually orthogonal unit vectors which are all orthogonal to the tangent space of M. Without loss of generality we assume $n(1) = \gamma$. Let $(g_{ij}) = (\langle \dot{\gamma}_i, \dot{\gamma}_j \rangle)$ denote the metric tensor of the manifold M and $ds = |g_{ij}|^{1/2}\, dt$ the volume element. Weyl (1939), display (10), showed that for all sufficiently small ϕ the volume $V(\phi)$ of the tube of geodesic radius ϕ about the manifold M without boundary is given by

$$V(\phi) = \int_M \left\{ \int \cdots \int_{\Sigma_2^{m-q}\tau_\nu^2 \le \tan^2\phi} \left| \delta_i^j - \sum_2^{m-q} \tau_\nu L_i^j(\nu) \right| \times \frac{d\tau_2 \cdots d\tau_{m-q}}{\left(1 + \Sigma_2^{m-q}\tau_\nu^2\right)^{m/2}} \right\} ds, \tag{4}$$

where (δ_i^j) is the identity matrix and $(L_i^j(\nu))$ is the matrix of the Weingarten map relative to the direction of $n(\nu)$ [cf. Millman and Parker (1977), page 125].

We also recall that the derivatives $\ddot{\gamma}_{ij} = \partial^2\gamma/\partial t_i\, \partial t_j$ satisfy Gauss's formulas

$$\ddot{\gamma}_{ij} = \sum_{k=1}^{q} \Gamma_{ij}^k \dot{\gamma}_k + \sum_{\nu=1}^{m-q} L_{ij}(\nu) n(\nu), \qquad i, j = 1, 2, \ldots, q, \tag{5}$$

where Γ_{ij}^k are Christoffel symbols and $L_{ij}(\nu)$ are coefficients of the second fundamental form relative to the direction $n(\nu)$, which are related to the Weingarten map by

$$L_{ij}(\nu) = \sum_{k=1}^{q} g_{ik} L_j^k(\nu) \tag{6}$$

[cf. Millman and Parker (1977), pages 104 and 125]. We can decompose U as

$$U = \sum_{i=1}^{q} a_i \dot{\gamma}_i + \sum_{\nu=1}^{m-q} \xi_\nu n(\nu). \tag{7}$$

From $\gamma = n(1)$, (7) and (5) we see that

$$Z = \xi_1$$

and

$$\ddot{Z}_{ij} = \sum_{k=1}^{q} \Gamma_{ij}^k \dot{Z}_k + \sum_{\nu=1}^{m-q} L_{ij}(\nu) \xi_\nu.$$

At a local maximum of Z we have $\dot{Z} = 0$, so by (6),

$$\ddot{Z}_{ij} = \sum_{\nu=1}^{m-q} \sum_{k=1}^{q} g_{ik} L_j^k(\nu) \xi_\nu$$

and hence

$$(8) \qquad |-\ddot{Z}_{ij}| = |g_{ij}| \cdot \left| -\sum_{\nu=1}^{m-q} L^i_j(\nu)\xi_\nu \right|.$$

From $\langle \gamma, \dot{\gamma}_i \rangle = 0$, $i = 1, \dots, q$, and (5) follow

$$L_{ij}(1) = \langle \ddot{\gamma}_{ij}, \gamma \rangle = -\langle \dot{\gamma}_i, \dot{\gamma}_j \rangle = -g_{ij}$$

and hence by (6) $L^i_j(1) = -\delta^i_j$, $i, j = 1, \dots, q$. Hence by (8) and (2) we have

$$(9) \quad E(N_w) = \int_T E\left\{ \left| \delta^i_j \xi_1 - \sum_{\nu=2}^{m-q} L^i_j(\nu)\xi_\nu \right|; Z > w, \dot{Z} = 0, \ddot{Z} < 0 \right\} |g_{ij}| dt,$$

where the expectation on the right is with respect to the joint distribution of the coefficients in (7).

If we express U as $U = (u_1, \dots, u_m)$ in terms of an orthonormal basis of $\mathbb{R}^m$, we can without loss of generality assume $\xi_\nu = u_\nu$, $\nu = 1, \dots, m-q$. Moreover, a calculation shows that the joint density of $(\dot{Z}_1, \dots, \dot{Z}_q)$ at $(0, \dots, 0)$ is $|g_{ij}|^{-1/2}$ times the joint density of $(u_{m-q+1}, \dots, u_m)$ at $(0, \dots, 0)$. Since the conditional distribution of $(u_1, \dots, u_{m-q})$ given $u_{m-q+1} = \cdots = u_m = 0$ is uniform on S^{m-q-1}, recalling that $Z = u_1$ and $ds = |g_{ij}|^{1/2}\, dt$, we obtain from (9) and some calculation the following result.

PROPOSITION 2. *For a manifold M without boundary,*

$$(10) \quad \begin{aligned} E(N_w) = \frac{\Gamma\left(\frac{m}{2}\right)}{\pi^{q/2}\Gamma\left(\frac{m-q}{2}\right)} \int_M E\Bigg\{ \left| \delta^i_j u_1^{(m-q)} - \sum_{\nu=2}^{m-q} L^i_j(\nu) u_\nu^{(m-q)} \right|; \\ u_1^{(m-q)} > w, \ddot{Z} < 0 \Bigg\} ds, \end{aligned}$$

where now expectation refers to the uniform distribution of

$$U^{(m-q)} = \left(u_1^{(m-q)}, \dots, u_{m-q}^{(m-q)}\right)$$

on S^{m-q-1} and

$$\ddot{Z} = \ddot{Z}(U^{(m-q)}) = -\sum_k g_{ik}\left(\delta^k_j u_1^{(m-q)} - \sum_\nu L^k_j(\nu) u_\nu^{(m-q)} \right).$$

Now assume that for all t,

$$(11) \qquad \{Z > w, \dot{Z} = 0\} \subset \{\ddot{Z} < 0\},$$

so the constraint on $\ddot{Z}$ in (10) is redundant. It is easy to see that the resulting expression is consistent with (4) as follows.

As is well known,

$$\left(u_1^{(m-q)}, \dots, u_{m-1}^{(m-q)}\right) =_{\mathscr{L}} (y_1, \dots, y_{m-q})/\|y\|,$$

where y_i's, $i = 1, \dots, m-q$, are independent $N(0,1)$ variables and $=_{\mathscr{L}}$ stands for equality in distribution. Hence

$$
\begin{aligned}
&E\left\{\left|\delta_j^i u_1^{(m-q)} - \sum_{\nu=2}^{m-q} L_j^i(\nu) u_\nu^{(m-q)}\right|; u_1^{(m-q)} > w\right\} \\
&\quad = E\left\{\left|\delta_j^i y_1/\|y\| - \sum_{\nu=2}^{m-q} L_j^i(\nu) y_\nu/\|y\|\right|; y_1/\|y\| > w\right\} \\
(12)\qquad &\quad = E\left\{\left|\delta_j^i - \sum_{\nu=2}^{m-q} L_j^i(\nu) y_\nu/y_1\right| (y_1/\|y\|)^q; y_1/\|y\| > w\right\} \\
&\quad = \frac{1}{(2\pi)^{(m-q)/2}} \int_{\{y_1 > \|y\| w\}} \left|\delta_j^i - \sum_{\nu=2}^{m-q} L_j^i(\nu) y_\nu/y_1\right| (y_1/\|y\|)^q \exp\{-\|y\|^2/2\}\, dy.
\end{aligned}
$$

Consider the transformation

$$\tau_1 = y_1, \tau_2 = y_2/y_1, \dots, \tau_{m-q} = y_{m-q}/y_1,$$

with

$$dy = \tau_1^{(m-q-1)}\, d\tau.$$

Let $\|\tau\|^2 = \sum_{\nu=2}^{m-q} \tau_\nu^2$, so $\|y\|^2 = y_1^2(1 + \|\tau\|^2)$. After some manipulation the integral in (12) becomes

$$2^{(m-q-2)/2}\Gamma\left(\frac{m-q}{2}\right) \int_{(\|\tau\|^2 < a^2)} \left|\delta_j^i - \sum_{\nu=2}^{m-q} L_j^i(\nu)\tau_\nu\right| (1 + \|\tau\|^2)^{-m/2}\, d\tau_2 \cdots d\tau_{m-q},$$

where $a^2 = w^{-2} - 1$. Therefore, under condition (11), $E(N_w)$ equals

$$
(13)\qquad \frac{\Gamma\left(\frac{m}{2}\right)}{2\pi^{m/2}} \int_M \left\{\int_{(\|\tau\|^2 < a^2)} \left|\delta_j^i - \sum_{\nu=2}^{m-q} L_j^i(\nu)\tau_\nu\right| \times \frac{1}{(1 + \|\tau\|^2)^{m/2}}\, d\tau_2 \cdots d\tau_{m-q}\right\} ds,
$$

which agrees with (3) and (4).

Display (13) says that under condition (11) the expected number of local maxima of Z at a height greater than w is the ratio of the volume of the tube about the manifold M of geodesic radius $\cos^{-1} w$ to the volume of the unit sphere. Condition (11) is satisfied whenever there is no overlap in the tube. In the case of no overlap, Weyl's formula gives the true volume. In particular, this is true whenever w is sufficiently close to 1, so the radius of the tube is sufficiently small.

For manifolds M with boundary ∂M, we must also consider those points of the tube for which the minimum distance to the manifold is attained at a point on the boundary. In general, the boundary of a q-dimensional manifold is a

$q - 1$-dimensional manifold which may itself have a $q - 2$-dimensional boundary. The general situation is quite complicated although essentially only one new remark is required to supplement the preceding calculations.

Suppose that M has a smooth *oriented* boundary, ∂M. That is, at each point $\gamma \in \partial M$ there exists a unit normal $n(2)$ to the tangent space of ∂M which lies in the tangent space of M and which points in the direction of the interior of M. If a point u in the tube around ∂M has a positive projection on the normal $n(2)$ at the nearest point $\gamma \in \partial M$, its contribution to the probability (1) has already been evaluated in Proposition 2. The new contribution to (1) is

$$P\{\langle \gamma, U\rangle > w \text{ and } \langle n(2), U\rangle \leq 0 \text{ for some } \gamma \in \partial M\}. \tag{14}$$

A result similar to Proposition 2 is easily obtained, but there is an additional constraint under the expectation sign in (10) corresponding to the condition $\langle n(2), U\rangle \leq 0$ in (14).

3. Comparisons with Weyl's result. This section contains two simple examples to illustrate that when the tube radius is large Weyl's formula can give a poor approximation to the desired tube volume in cases where the expected number of local maxima yields the correct result.

EXAMPLE 1 (Sphere in $\mathbb{R}^m$). For simplicity, we first consider the case where our manifold is the unit sphere S^{m-1} in $\mathbb{R}^m$ and the tube is the set of all points within Euclidean distance w of the sphere. A random point U is uniformly distributed in a large ball of radius a containing the manifold of interest, and the random field X is the Euclidean distance from U to a point γ in the manifold. Of course, it is trivial to determine directly the volume of a tube about S^{m-1}. The point of the example is to show that our upper bound always yields the exact volume, whereas for large tube radii Weyl's formula overestimates the tube volume when m is odd and underestimates it when m is even. We begin by considering an arbitrary hypersurface and later specialize to the unit sphere.

Let $X = \|\gamma - U\|^2$. We assume that $1 + w < a$, so the tube of Euclidean radius w about the hypersurface M is completely contained in the ball of radius a centered at the origin. If N denotes the number of local *minima* of X where $X < w^2$, the analog of Proposition 1 is

$$EN = \int_T E\big(|\ddot{X}|;\, X < w^2,\, \dot{X} = 0,\, \ddot{X} > 0\big)\, dt.$$

From $\dot{X}_i = 2\langle \dot{\gamma}_i, \gamma - U\rangle$, $\ddot{X}_{ij} = 2\langle \ddot{\gamma}_{ij}, \gamma - U\rangle + 2g_{ij}$, we see that at points γ where $\dot{X} = 0$, U lies in the direction normal to M, and from Gauss' formulas (5)

$$\ddot{X}_{ij} = 2g_{ij} + 2L_{ij}\langle n, \gamma - U\rangle,$$

where n is normal to the manifold M at γ.

For the special case where our manifold is the unit sphere, $n = \gamma$ and hence $L_{ij} = \langle n, \ddot{\gamma}_{ij} \rangle = -\langle \dot{\gamma}_i, \dot{\gamma}_j \rangle = -g_{ij}$. Simple algebra now shows that for the special case of interest

$$EN = \int_T E\Big(\langle \gamma, U \rangle^{m-1}; (1-w)^+ < \langle \gamma, U \rangle < 1 + w, \tag{15}$$
$$\langle \dot{\gamma}_i, U \rangle = 0, \text{ for all } i\Big)|g_{ij}|\, dt.$$

The joint density of $\langle \gamma, U \rangle, \langle \dot{\gamma}_i, U \rangle$, $i = 1, \dots, m-1$, at $(x, 0, \dots, 0)$ is $[|g_{ij}|^{1/2} a^m \Omega_m]^{-1}$, where Ω_m is the volume of the unit ball in $\mathbb{R}^m$. Hence

$$EN = \Big\{(1+w)^m - \big[(1-w)^+\big]^m\Big\}/a^m,$$

which is the volume of the tube divided by the volume of the ball of radius a. It is easy to see that $P[\min X \le w^2] = E(N_w)$. Weyl's formula gives a very similar result, except that there is no positive part in (15) or the final expression. When $w < 1$, it gives the true probability; otherwise it is a lower or upper bound for the probability depending on whether m is even or odd, and the error made by Weyl's formula can be substantial.

This example has an instructive geometric interpretation. Assume $1 < w < 2$. Suppose that U is inside the unit ball of radius $w - 1$ centered at the origin and $X(\gamma)$ is the squared distance from U to an arbitrary point γ on the unit sphere. As γ moves on the unit sphere, X attains its extrema at two points. An elementary argument shows that we can without loss of generality assume $U = (u_1, 0, \dots, 0)$, where $0 < u_1 < w - 1$. Since the unit sphere satisfies the equation $x_1^2 + \cdots + x_m^2 = 1$, the squared distance from $\gamma = (x_1, \dots, x_m)$ to U is

$$(x_1 - u_1)^2 + \cdots + x_m^2 = 1 - 2u_1 x_1 + u_1^2,$$

which attains its extrema, one minimum and one maximum, when $|x_1| = 1$. Our formula includes the minimum and excludes the maximum to arrive at the correct probability. Depending on whether m is even or odd, the expression given by Weyl's formula is less or larger than the true probability. At the maximum the $(m-1) \times (m-1)$ determinant of Weyl's formula has entries $\langle \gamma, U \rangle < 0$ along the diagonal and 0 elsewhere, and hence is negative when m is even and positive otherwise. A picture in two dimensions is illuminating. See also (15).

EXAMPLE 2 (Cylinder). Consider the sphere S^3 in $\mathbb{R}^4$ and a cut cylinder M in S^3,

$$M = \Big\{\gamma(\theta, \phi) = \Big(h \cos\theta \cos\phi, h\cos\theta\sin\phi, h \sin\theta, \sqrt{1-h^2}\Big):$$
$$\theta \in (0, \pi), \phi \in (0, 2\pi)\Big\},$$

where h is fixed in $(-1, 1)$. This surface is a natural two-dimensional version of an example of Knowles, Siegmund and Zhang (1991).

We have

$$
\begin{aligned}
Z &= h\cos\theta(\cos\phi u_1 + \sin\phi u_2) + h\sin\theta u_3 + \sqrt{1-h^2}\,u_4,\\
\dot{Z}_\theta &= -h\sin\theta(\cos\phi u_1 + \sin\phi u_2) + h\cos\theta u_3,\\
\dot{Z}_\phi &= h\cos\theta(-\sin\phi u_1 + \cos\phi u_2),\\
\ddot{Z}_{\theta\theta} &= -h\cos\theta(\cos\phi u_1 + \sin\phi u_2) - h\sin\theta u_3,\\
\ddot{Z}_{\theta\phi} &= -h\sin\theta(-\sin\phi u_1 + \cos\phi u_2),\\
\ddot{Z}_{\phi\phi} &= -h\cos\theta(\cos\phi u_1 + \sin\phi u_2).
\end{aligned}
$$

Conditional on $\dot{Z}_\theta = 0$, $\dot{Z}_\phi = 0$, we have $\ddot{Z}_{\theta\phi} = 0$ and

$$
\begin{aligned}
\ddot{Z}_{\phi\phi} &= -\frac{h\cos\theta}{\cos\phi}u_1,\\
\ddot{Z}_{\theta\theta} &= \ddot{Z}_{\phi\phi} - h\sin\theta u_3\\
&= \ddot{Z}_{\phi\phi} = \frac{\sin\theta}{\cos\theta}\left(\frac{\sin\theta}{\cos\theta}\ddot{Z}_{\phi\phi}\right)\\
&= \frac{\ddot{Z}_{\phi\phi}}{\cos^2\theta} = -\frac{h}{\cos\phi\cos\theta}u_1.
\end{aligned}
$$

Then the expression given by Weyl's formula is

$$
CE\left(\frac{h^2}{\cos^2\phi}u_1^2;\ \frac{h}{\cos\phi\cos\theta}u_1 + \sqrt{1-h^2}\,u_4 > w\right), \tag{16}
$$

where C is some constant playing no role in our illustration, and the expectation of the number of local maxima is

$$
CE\left(\frac{h^2}{\cos^2\phi}u_1^2;\ \frac{h}{\cos\phi\cos\theta}u_1 + \sqrt{1-h^2}\,u_4 > w,\ \frac{h}{\cos\phi\cos\theta}u_1 > 0\right). \tag{17}
$$

We see that (16) $\geq$ (17). The equality holds if and only if $1 - h^2 < w^2$; that is, there is no overlap in the tube.

4. Numerical evaluation. In Section 2 we showed that Weyl's formula is exact if condition (11) is satisfied. However, in statistical applications it is not feasible to verify this condition; and examples in the previous section, although artificial, show that neglecting the condition can lead to arbitrarily poor approximations to the true probability. For the special case of a two-dimensional manifold, Zhang (1991) has evaluated numerically the expression given in Proposition 2 for $E(N_w)$. In Section 5 we use this evaluation to compare our upper bound with Weyl's approximation in a number of statistical examples.

The evaluation of $E(N_w)$ involves a lengthy calculation, which is divided into a number of different cases. The function γ and its derivatives are represented in terms of an orthonormal basis of $\mathbb{R}^m$. Repeated use is made of the fact that if $U^{(n)} = (u_1^{(n)}, \dots, u_n^{(n)})$ is uniformly distributed on S^{n-1}, then for $p < n$, the joint density of $(u_1^{(n)}, \dots, u_p^{(n)})$ is given by

$$\frac{\Gamma\left(\dfrac{n}{2}\right)}{\pi^{p/2}\Gamma\left(\dfrac{n-p}{2}\right)}\left(1 - \sum_1^p u_i^2\right)^{(n-p-2)/2}. \tag{18}$$

The details are quite complicated and are not given here. The final expression involves a four-dimensional numerical integration. This should be contrasted with Weyl's formula, which always involves an integration of the dimension of the manifold—2 in this case. In one dimension Naiman's (1986) upper bound is no more difficult to evaluate than Hotelling's (1939) equality, but we have been unable to achieve a similar result in higher dimensions. The added complication involved in evaluation of $E(N_w)$ comes from the constraint that $\ddot{Z}$ be negative definite. We do not know for a higher-dimensional manifold how many additional dimensions of numerical integration are required, but the number appears to grow rapidly with the dimension. We believe it is possible to carry out the numerical integration for a manifold of three dimensions, but four or more dimensions pose quite substantial challenges.

In the following numerical examples, we often do the calculations with and without the constraint on $\ddot{Z}$ in order to obtain some empirical evidence concerning the possibility of using the simpler evaluation as an approximation of the more complicated one. For the probability discussed so far, this approximation just brings us back to Weyl's formula. However, for the problem of confidence regions discussed in Section 6, the corresponding approximation does not have a simple interpretation and apparently cannot be obtained by local methods of differential geometry.

5. Examples and simulation. In this section we consider statistical applications of the results stated in Section 2. Since complete evaluation of the expressions given there is onerous, we shall be interested in the extent to which neglecting some aspects of the calculation leads to useful approximations.

Assume that we have observations (x_i, y_i), $i = 1, \dots, m$, from the regression model

$$y_i = X_i\alpha + \beta f_i(x, \theta) + \varepsilon_i, \tag{19}$$

where X_i is a vector which is independent of unknown parameters, f_i is a nonlinear function of θ and ε_i are independent and identically distributed as $N(0, \sigma^2)$ with unknown σ. Suppose we want to test H_0: $\beta = 0$. For this purpose we can assume without loss of generality that the linear term X_i does not appear in (19) [cf. Siegmund and Zhang (1991)]. Then the p-value of the likelihood ratio test for H_0 is

$$P\left[\sup_{\theta}|\langle\gamma(\theta), U^{(m)}\rangle| > w\right], \tag{20}$$

where, under H_0, $U^{(m)}$ is uniformly distributed on S^{m-1}, γ is the unit vector $f/\|f\|$, w is the observed value of $|\langle\gamma(\hat{\theta}), y\rangle|/\|y\|$ and $\hat{\theta}$ is the maximum likelihood estimator of θ.

EXAMPLE 3 (Testing for a harmonic). We begin with an example from Knowles and Siegmund (1989). The model is

$$y_i = \beta\cos(\mu x_i + \omega) + \varepsilon_i,$$

where for simplicity $x_i = i - (m+1)/2$, $i = 1/, \dots, m$.

For various values of m and w, Knowles and Siegmund computed p-values from Monte Carlo and from their Corollary 3, derived from Weyl's formula. In some cases they were able to use the special structure of the model to show that the expression given by Weyl's method is a genuine upper bound. In other cases they were unable to show this, and evidence from their Monte Carlo experiment was ambiguous.

The method of this paper gives numerical results which are virtually indistinguishable from those of Knowles and Siegmund in all the cases they considered. Thus, insofar, as this example is concerned, the theoretical possibility that Weyl's method may be anticonservative seems not to be realized. See Zhang (1991) for the actual numerical values.

EXAMPLE 4 (Broken-plane mode). The "broken-plane" model is a natural generalization of the broken-line regression model discussed by a number of authors. A particularly interesting data set involving the lifetime of plastic pipes as a function of stress and temperature is mentioned by van de Geer (1988) and has been provided to the authors by Richard Gill.

The model is piecewise linear and because of the linearity seems to be a particularly simple example for our general theory; but by virtue of the lack of differentiability of the regression surface, our results do not apply directly. To understand these problems in the simplest context, it is helpful to consider the tube of Euclidean radius w about a broken line γ in the plane (cf. Figure 1) and compare the approach of this paper with the related but intrinsically one-dimensional level crossing theory of Johnstone and Siegmund (1989). [See Siegmund and Zhang (1991) for an application of that idea to broken-line regression and Naiman (1986) for a similar picture described analytically.] Using the expected number of downcrossings (cf. Example 1), one obtains the same upper bound to the volume of the tube either by a direct argument applied to the broken line or by approximating the broken line uniformly by a sequence of smooth curves, since for each point in the interior of the tube, that is, whose distance to the broken line is less than the tube radius, the number of downcrossings is the same for all smooth approximating curves that are sufficiently close to the broken line. The expected number of downcrossings

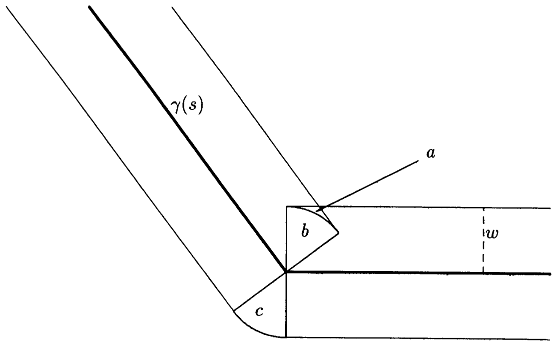

FIG. 1. *Broken line in the plane.*

gives a strict upper bound for the volume of the tube in Figure 1 because points in the figure designated with the letter a are counted twice, while points in the areas b and c are counted once, as they should be. The situation regarding the expected number of local minima is more complex. If we approximate the broken line by a sequence of smooth curves, the points in areas a and b are both counted twice, while those in area c are counted once, to give a less accurate bound. On the other hand, if we apply the local minima calculation directly to the broken line, points U in areas a and b are counted twice, while those in c are not counted at all, since there the derivative of the process $\|\gamma - U\|^2$ jumps across 0 without ever being equal to 0. Since the areas b and c are equal, the expected number of local minima exactly equals the expected number of downcrossings, so the two methods give the same bound on the tube volume. Note, however, that this last argument would not work for an arbitrary continuous curve with a discontinuous derivative, so to obtain a guaranteed upper bound for the tube volume by using local minima, one would have to calculate the limit associated with a sequence of smooth approximations.

For broken-plane regression, there is no notion of the expected number of upcrossings, but we can easily calculate the expected number of local maxima. In what follows we have employed the symmetry considerations of the preceding paragraph to obtain the better of the two bounds. The difference between these bounds can be quite substantial if the number of breaks in the regression surface is large.

We consider the simplest possible broken-plane model

$$y = \beta(x_1 - \theta_1 x_2 - \theta_2)^+. \tag{21}$$

Suppose that x_1 takes values of $1, \dots, 6$; to each value of x_1 there corresponds a series of x_2 taking values of $1, \dots, 7$. So there are 42 design points of (x_1, x_2). Let

$$\Lambda = \sup_{\theta_1, \theta_2} \langle \gamma(\theta_1, \theta_2), U^{(42)} \rangle^2. \tag{22}$$

TABLE 1
Comparison of p-values: $P[\Lambda > w^2]$

w^2	Empirical value	Approximated value
0.285	0.010	0.010
0.249	0.025	0.025
0.221	0.050	0.051
0.203	0.075	0.079
0.191	0.100	0.108
0.180	0.125	0.140
0.171	0.150	0.174

Finding $\hat{\theta}$ for (19) is usually difficult; fortunately, the special structure of the model (21) allows us to obtain the maximum likelihood estimates quite efficiently. See the algorithms described in van de Geer (1988) and Zhang (1991). Thus it is feasible to approximate the empirical distribution of Λ and hence the p-value (20) by Monte Carlo.

Our Monte Carlo experiment proceeds as follows. First, we independently draw 42 observations of $y_1, \dots, y_{42}$ from $N(0, 1)$; we fit the sample data to the model (21) using the algorithm suggested in Section 7.2 of Zhang (1991); we find the regression residual sum of squares, denoted by R, and convert R into Λ via the relationship $\Lambda = 1 - R/\|y\|^2$. Now we repeat this procedure 20,000 times to obtain the empirical distribution of Λ. See Table 1 for selected quantiles of the tail of the distribution.

On the other hand, an approximation can be derived for the distribution of Λ by our local maxima results. It is interesting to point out that because of the piecewise linearity of the manifold, the bound obtained by the local maxima argument depends only on the area of the manifold $M = \{\gamma(\theta_1, \theta_2)\}$, the length of its boundary ∂M and the sum of the exterior angles ϕ_i. The numerical computation of these quantities is easy because every edge on the manifold M is a segment of a big circle. Some numerical results are given in Table 1. The accuracy of our bound, especially in the upper tail, seems more than sufficiently accurate for practice.

We have also applied this method to obtain a p-value for the test of no break point in the regression surface for the data provided by Richard Gill (see Figure 3), concerning survival time of plastic pipes as a function of temperature and stress. The result was an attained significance less than 10^{-5}.

6. Confidence regions. In Section 5 we observed that testing the hypothesis $\beta = 0$ for the model (19) leads to the problem of evaluating the probability (20). Now we are interested in finding confidence regions for θ, or more generally joint regions for θ and β and/or some components of α for the model (19). Conceptually, our method is exactly the same as that discussed by Knowles, Siegmund and Zhang (1991) for models involving only one nonlinear parameter, but the required probability calculations are more complicated. See also Siegmund and Zhang (1991). For simplicity, we again assume that the linear term involving X_i does not appear in (19). Our $1 - \alpha$ confidence region for θ is the set of all θ_0 such that

$$P_{\theta_0}\left[\sup_\theta \langle \gamma(\theta), U^{(m)}\rangle^2 > w^2 | \langle \gamma(\theta_0), U^{(m)}\rangle = z\right] > \alpha, \tag{23}$$

where w and z are the observed values of $\max_\theta |\langle \gamma(\theta), y/\|y\|\rangle|$ and $\langle \gamma(\theta_0), y/\|y\|\rangle$, respectively. For a manifold without boundary, we bound from above the probability on the left-hand side of (23) by

$$E^z[N_w(M)] = \int \mathbf{E}^z\left[|\ddot{Z}|; Z > w, \dot{Z} = 0, \ddot{Z} < 0\right] d\theta, \tag{24}$$

where E^z means conditional on $\langle \gamma(\theta_0), U^{(m)}\rangle = z$.

Knowles, Siegmund and Zhang (1991) give a geometric interpretation for (23), although there does not seem to be a (local) geometric method for calculating it. The numerical evaluation of (24) is similar to that of (10) though more complicated. We omit the tedious calculations and refer to Zhang (1991) for details.

There are a number of other methods for obtaining confidence regions in nonlinear regression [cf. Bates and Watts (1988), Chapters 6 and 7, or Seber and Wild (1989), Chapters 3 and 5]. One of these, the approximate F or likelihood method, is based on the assumption that an appropriate version of the likelihood ratio statistic has an F distribution, as it would in the case of a linear model. This method seems to perform very well, better in fact than the theory behind it leads one to suspect, although Knowles, Siegmund and Zhang (1991) show that it can behave poorly when the confidence region contains points close to the boundary of the parameter space.

EXAMPLE 5 (Hunt's ryegrass data). We consider data published by Hunt (1971) and also analyzed by Cook and Goldberg (1986). Hunt originally suggested use of the logistic model

$$y = \frac{\beta}{1 + \exp(\theta_1 + \theta_2 x)} \tag{25}$$

for the data. However, Cook and Goldberg fit the data with an asymptotic growth model

$$y = \beta_0 + \beta \exp(-\theta x). \tag{26}$$

Our interest in these data was prompted by the unusual shape of the confidence regions obtained by Cook and Goldberg. This shape suggests large curvature, where methods like the approximate F method, which are justified by local, linear approximations, may break down. This should then be a good test case for our method, which can only err in the direction of conservatism, and which, based on the empirical evidence in Knowles, Siegmund and Zhang (1991), we believe to be essentially exact under a wide variety of conditions. A plot of the data shows clearly the S shape associated with the logistic model

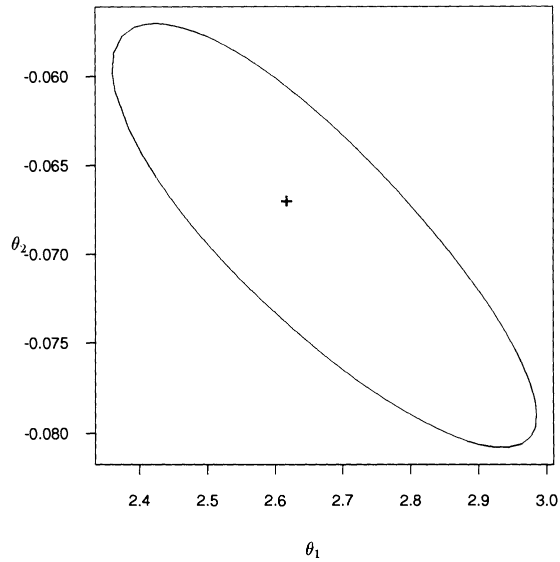

FIG. 2. *95% joint confidence region for ryegrass data. The solid region was obtained by the approximate F method; the plus sign is the location of the MLE.*

and incompatible with the asymptotic growth model. Indeed, the two components of the confidence region obtained by Cook and Goldberg indicate the schizophrenia of the asymptotic growth model in trying to choose parameter values which either ignore the convex curvature at the left-hand end of the regression function while modeling the concave curvature at the right-hand end, or vice versa. Hence we have used the logistic model (25).

Since evaluation of (24) is time consuming and unlike the significance level in the preceding sections must be done repeatedly, Knowles, Siegmund and Zhang (1991) suggest the possibility of calculating the much simpler approximate F confidence region and using their method as a diagnostic to check the accuracy of that region and if appropriate to correct it. In Figure 2 we have obtained a 95% approximate F joint confidence region for (θ_1, θ_2). It is approximately elliptical in shape, which suggests that it should have close to its nominal coverage probability. As a check, using (24) we computed upper bounds for conditional p-values at 15 points along the boundary of the 95% approximate F region. The results are given in Table 2. It is apparent that the approximate F method has given a very accurate region in this case.

Since one of the most onerous aspects of evaluating (24) numerically is the constraint that $\ddot{Z}$ be negative definite, we performed the same calculation omitting this constraint. The calculation was much faster and yielded the same numerical results. We saw in Section 2 that in the hypothesis-testing context, eliminating this constraint gives us Weyl's formula; but for the present problem we do not know any comparable interpretation.

TABLE 2
Tail probability along boundary

θ_1	θ_2	**Upper bound**
2.362	−0.059	0.0498
2.460	−0.057	0.0496
2.560	−0.059	0.0499
2.660	−0.062	0.0498
2.760	−0.066	0.0495
2.860	−0.070	0.0499
2.960	−0.076	0.0498
2.985	−0.079	0.0497
2.964	−0.081	0.0497
2.864	−0.080	0.0497
2.764	−0.078	0.0497
2.664	−0.075	0.0497
2.564	−0.072	0.0497
2.464	−0.068	0.0498
2.364	−0.061	0.0496

Although the model (27) seems inappropriate, we also assumed that model and calculated a joint confidence region for (θ, β), which was virtually indistinguishable from the disconnected approximate F region of Cook and Goldberg (1986).

EXAMPLE 6 (Broken-plane model). Now, we study the data on plastic pipes provided by Richard Gill. The survival time of a pipe, that is, the time until the pipe bursts, is believed to depend on the ambient temperature and on the stress applied to it. The data are presented in Figure 3. The logarithm of the stress (x_2) in MN/m^2 (meganewtons/square meter) and the logarithm of the survival time y in hours are plotted along the horizontal and vertical axes, respectively. There are only four values of temperature: 20°C, 40°C, 60°C and 80°C. The reciprocal of the temperature is chosen as x_1. We use the broken-plane model

$$y = \alpha_1 + \theta_1 x_1 + \beta_1 x_2 + \beta_2 (x_2 - \theta_2 x_1 - \alpha_2)^+. \tag{27}$$

The predicted change line is $x_2 = -0.0043 + 1.14 x_1$; and for each fixed temperature of 20°C, 40°C, 60°C and 80°C, the fits are shown in Figure 3. As we indicated previously, the p-value of the likelihood ratio test for the hypothesis $\beta_2 = 0$ is less than 10^{-5}. Here we compute the 95% approximate F and conditional likelihood ratio confidence regions for (θ_2, α_2). Figure 4 reports the results. Both regions are in close agreement and are elliptical. The small variations of the estimated coefficients of the change line suggests that the model is reasonable for these data and fulfills the goal of allowing us to estimate the break point in the line at 20°C from the data obtained at higher temperatures.

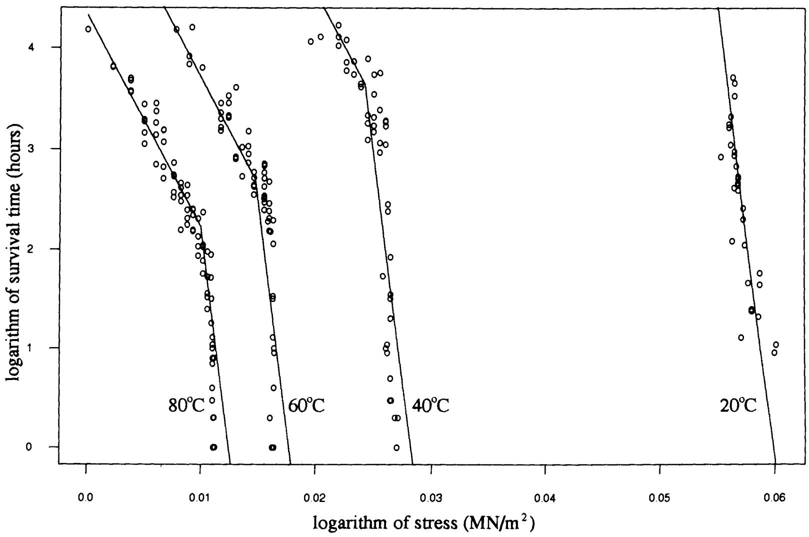

FIG. 3. *Data for plastic pipe.*

Because of the nondifferentiability of the manifold in this model, the computations leading to Figure 4 present some special problems, which were mentioned in Example 4. For the evaluation of the confidence region in Figure 4, we did not introduce a smooth approximating manifold, but rather took advantage of the special structure of the broken-plane manifold to simplify the calculation. See Zhang (1991) for a detailed derivation. However, unlike the

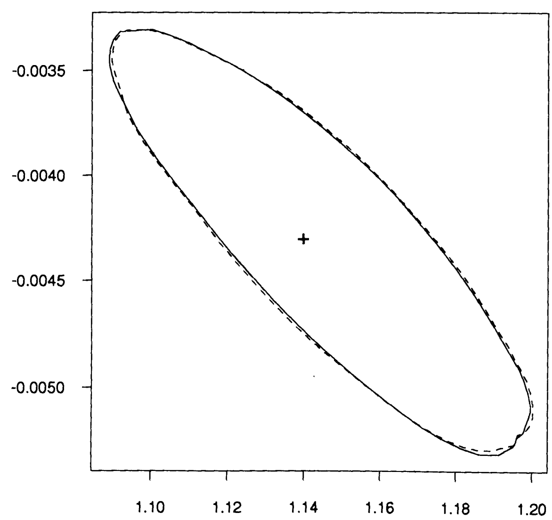

FIG. 4. *95% confidence region for* (θ_2, α_2). *The solid and dashed regions are the AF and CLR ones, respectively.*

hypothesis-testing situation in Example 4, because of the conditioning involved, we cannot argue by symmetry that the result we obtain is a genuine upper bound, although it should be a reasonable approximation. It is also plausible, however, that our calculation is not numerically stable and by doing the calculation somewhat differently we might obtain different numerical results.

The manifold for this model has a boundary, so in principle there should be a boundary term to supplement (24) just as there was the boundary correction to the basic expression (10). For these data, inclusion of an appropriate boundary term has no discernible effect on the confidence region. Our experience indicates that this is quite often the case, although it appears to be precisely those cases where the boundary term is important that the approximate F method, predicated on an approximately linear model with no boundary, may break down [cf. Knowles, Siegmund and Zhang (1991) and Zhang (1991)].

REFERENCES

ADLER, R. J. (1981). *The Geometry of Random Fields*. Wiley, New York.

BATES, D. M. and WATTS, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. Wiley, New York.

BELYAYEV, Y. K. (1972). Point process and first passage problem. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **3**. Univ. California Press, Berkeley.

COOK, R. D. and GOLDBERG, M. L. (1986). Curvatures for parameter subsets in nonlinear regression. *Ann. Statist.* **14** 1399–1418.

ESTERMANN, T. (1926). Über Carathéodorys und Minkowskis Verallgemeinerungen des Längenbegriffs. *Abh. Math. Sem. Univ. Hamburg* **4** 73–116.

GRAY, A. (1982). *Tubes*. Addison-Wesley, Reading, MA.

HOTELLING, H. (1939). Tubes and spheres in n-space and a class of statistical problems. *Amer. J. Math.* **61** 440–460.

HUNT, W. F. (1971). The influence of leaf death on the rate of accumulation of green herbage during pasture regrowth. *Ecology* **7** 41–50.

JOHANSEN, S. and JOHNSTONE, I. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18** 652–684.

JOHNSTONE, I. and SIEGMUND, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality. *Ann. Statist.* **17** 184–194.

KNOWLES, M. (1987). Simultaneous confidence bands for random functions. Ph.D. dissertation, Stanford Univ.

KNOWLES, M. and SIEGMUND, D. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression. *Internat. Statist. Rev.* **57** 205–220.

KNOWLES, M., SIEGMUND, D. and ZHANG, H. (1991). Confidence regions in semilinear regression. *Biometrika* **78** 15–31.

LEADBETTER, M., LINDGREN, G. and ROOTZEN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

MILLMAN, R. S. and PARKER, G. D. (1977). *Elements of Differential Geometry*. Prentice-Hall, Englewood Cliffs, NJ.

NAIMAN, D. Q. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.* **14** 896–1906.

NAIMAN, D. Q. (1987). Simultaneous confidence bands in multiple regression using predictor variable constraints. *J. Amer. Statist. Assoc.* **82** 214–219.

NAIMAN, D. Q. (1990). Volumes of tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Statist.* **18** 685–716.

SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

SIEGUMND, D. and ZHANG, H. (1991). Confidence regions in broken line regression. Technical Report, Stanford Univ.

SMALE, S. (1981). The fundamental theorem of algebra and complexity theory. *Bull. Amer. Math. Soc.* **4** 1–36.

SUN, J. (1989). *P*-values in projection pursuit. Ph.D. dissertation, Dept. of Statistics, Stanford Univ.

VAN DE GEER, S. A. (1988). Regression analysis and empirical process, Ph.D. dissertation, Univ. Leiden.

WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–472.

ZHANG, H. (1991). Confidence regions in nonlinear regression and geometry, Ph.D. dissertation, Dept. Statistics, Stanford Univ.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF EPIDEMIOLOGY AND PUBLIC HEALTH
YALE UNIVERSITY SCHOOL OF MEDICINE
60 COLLEGE STREET
NEW HAVEN, CONNECTICUT 06510