The Annals of Statistics
1990, Vol. 18, No. 2, 652–684

HOTELLING'S THEOREM ON THE VOLUME OF TUBES: SOME ILLUSTRATIONS IN SIMULTANEOUS INFERENCE AND DATA ANALYSIS

BY SØREN JOHANSEN AND IAIN M. JOHNSTONE

University of Copenhagen and Stanford University

We illustrate with contemporary examples Hotelling's geometric approach to simultaneous probability calculations. Hotelling reduces the evaluation of certain normal theory significance probabilities to finding the volume of a tube about a curve in a hypersphere, and shows that this volume is often exactly given by length times cross-sectional area. We review Hotelling's result together with some recent complements, and then use the approach to set simultaneous prediction regions for some data from gait analysis, to study Andrews' plots in multivariate data analysis, and to construct significance tests for projection pursuit regression. A by-product is a numerical criterion for tube self-overlap, relevant, for example, to uniqueness of certain nonlinear least squares estimates.

1. Introduction and summary. Harold Hotelling (1939) proved that the volume of a tube about a smooth closed curve in Euclidean space or a hypersphere exactly equals the length of the curve multiplied by its cross-sectional area, so long as the tube does not overlap itself. Although Hotelling's intended application lay in hypothesis tests, his striking result has had far greater influence in geometry, largely through its far-reaching extension in the companion paper by Weyl (1939). Within statistics, the methods and motivation of Hotelling's paper have stimulated work on "volume tests" of significance [Diaconis and Efron (1985) and references therein] and on the order of singularities in sampling distributions of t, F and r [e.g., Bradley (1952), Siddiqui (1958) and Mulholland (1965, 1970)]. However, the explicit tube result itself appears to have been little used by statisticians.

Our purpose in this partly expository paper is to set forth some contemporary examples in simultaneous inference and data analysis where we have found the tubes method to be helpful. With the exception of Section 6, the focus is on situations involving *curves* in spheres. Knowles and Siegmund (1989) give an exposition with statistical extensions and applications of Weyl's result for *surfaces* in spheres, concentrating on the two-dimensional case.

The expository section, Section 2, reviews Hotelling's result and an important complement of Naiman (1986). We then describe some classes of applications: testing for a nonlinear parameter, simultaneous confidence and prediction bands and then how these are reduced to the Hotelling–Naiman

Received May 1985; revised March 1989.

AMS 1980 *subject classifications*. Primary 62E15; secondary 62J15, 62F25, 62J02.

Key words and phrases. Volume of tubes, simultaneous inference, confidence bands, prediction bands, gait analysis, Andrews' plots, projection pursuit regression.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to
The Annals of Statistics.
www.jstor.org

results. Some simple upper bounds follow and finally a simple form of Weyl's result tailored for use in Section 6.

The succeeding sections present diverse variations on Hotelling's theme. The unifying thread is the data-analytic or probabilistic interpretation of the distance of a unit vector to a curve (or surface) in a higher-dimensional sphere.

Section 3 tests Hotelling's theory on some gait data analyzed at length by Olshen, Biden, Wyatt and Sutherland (1989). These authors use bootstrap methods to develop simultaneous prediction regions for diagnosing abnormal gait in young children. Hotelling's approach can be readily adapted, and in all cases tried, a simple analytic approximation reproduces the tail of the bootstrap distribution tolerably well.

Hotelling's formula is only exact for tubes of sufficiently small radius that no self-overlap occurs. Otherwise, as Naiman (1986) shows, the method leads only to an upper bound. For applications to simultaneous confidence and prediction regions, Hotelling's formula must be averaged over all tube radii, so that overlap is unavoidable in these cases. What, then, is the largest radius for which the formula is exact? One cannot expect to say much theoretically in general, but for any given application, Section 4 derives an easily computed bivariate function whose minimum yields the desired critical radius.

The overlap formula is illustrated in the course of a discussion of Andrews' (1972) plots in Section 5. The Andrews plot represents points in high-dimensional space $\mathbb{R}^d$ by graphs of trigonometric polynomials. We regard the plot as a simple-to-analyze forerunner of more recent "grand tour" methods for projection pursuit explorations of data and ask what fraction of possible projections are seen, assuming a given "squint angle" for the data analyst. The squint angle determines the radius of the tube whose volume we compute via the Hotelling formula. A simulation experiment shows the Hotelling formula to be useful even for some radii considerably above the threshold for overlap, especially in higher dimensions. There is also a simple expression for the distance of the furthest (unseen) projection from the curve in odd dimensions.

The final section presents a prototype for application of the tubes approach to construction of significance tests in projection pursuit. Projection pursuit methods involve extensive "data dredging" in the search for interesting views, so it is important to be able to discriminate real from spurious structure. The search over directions usually involves the maximization of a "projection index," which can sometimes be interpreted in terms of the distance of the data from a surface. As a concrete example, we take a version of projection pursuit regression based on orthogonal polynomials and Gaussian independent variables. Weyl's extension of Hotelling's result is needed since the high dimensions of projection pursuit entail replacing curves by manifolds of higher dimension (but relatively simple structure). We give an approximate formula for the significance level of a simple test of the null hypothesis of zero regression (a limit theorem gives the precise statement). While this particular example may not be suited to application, similar methods are being used by

Sun (1989) in the context of Friedman's (1987) practical implementation of exploratory projection pursuit.

There are a number of other applications of tube methods that we do not discuss here. Knowles (1987) shows how they may be used to derive some old and new bounds for the distribution of suprema of smooth Gaussian processes and fields. Simultaneous posterior credible regions can be derived for conjugate priors in the linear model settings described in Section 2.

In combination, these and other examples to be reported elsewhere, suggest that Hotelling's method remains relevant because it provides a heuristic for illuminating an increasing number of techniques that involve *maximally selected inference*. These methods often involve the selection of a projection, or variable, or view α to maximize an objective $T_n^2(\alpha)$. Under an appropriate null model, the distribution of $T_n^2(\alpha)$ for fixed α can be related to a spherical cap $C(\alpha)$ and a significance level or P-value is tied to the volume of $C(\alpha)$. Thus, calibration of $\max_\alpha T_n^2(\alpha)$ involves the volume of $\bigcup_\alpha C(\alpha)$, which is often exactly or approximately a tube (or tubes) in the sphere. Depending on whether the setting is parametric, or nonparametric, the heuristic can yield either careful numerical approximations [as, e.g., in Knowles and Siegmund (1989)] or rough guidelines, as in the projection pursuit setting of Section 6.

Finally, we establish some notation. $S^{d-1} = \{x \in \mathbb{R}^d\colon |x| = 1\}$ denotes the $(d-1)$-dimensional unit sphere embedded in $\mathbb{R}^d$. It has $(d-1)$-dimensional volume ("surface area") $\omega_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$. Here $\Gamma(r)$ denotes the gamma function and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. The d-dimensional unit ball $\{x \in \mathbb{R}^d\colon |x| \le 1\}$ has d-dimensional volume $\Omega_d = \pi^{d/2}/\Gamma(d/2 + 1)$. These quantities are related via $\Omega_d = \omega_{d-1}/d$.

2. Hotelling's formula, Naiman's bound and principal applications.

A. *Hotelling's formula and Naiman's bound for tubes in spheres.* Let $I = [a, b] \subset \mathbb{R}$ be a closed interval and $\gamma\colon I \to S^{d-1}$ a *regular* [continuously differentiable with nowhere vanishing derivative $\dot\gamma(t) = d\gamma/dt$] curve lying in the unit sphere $S^{d-1} \subset \mathbb{R}^d$. We demand that γ have no self-intersections, excepting possibly $\gamma(a) = \gamma(b)$, in which case we further require that $\dot\gamma(a+) = \dot\gamma(b-)$ and γ is said to be *closed*. Abusing notation, we also denote the image of γ by γ. The *length* of γ is $|\gamma| = \int_I|\dot\gamma(t)|\,dt$. The distance of a point $u \in S^{d-1}$ to the curve γ is the distance to the closest point of γ: $d^2(u,\gamma) = \inf_t |u - \gamma(t)|^2 = 2(1 - \sup_t u'\gamma(t))$. Define the *tube* of angular (geodesic) radius θ about γ in S^{d-1} by (see Figure 1)

$$\gamma^\theta = \{u \in S^{d-1}\colon \sup_t u'\gamma(t) \ge \cos\theta\}$$

$$= \{u \in S^{d-1}\colon d(u,\gamma) \le (2(1-w))^{1/2}\}, \qquad w = \cos\theta.$$

Let $V(\gamma^\theta)$ denote the volume (in S^{d-1}) of the tube γ^θ.

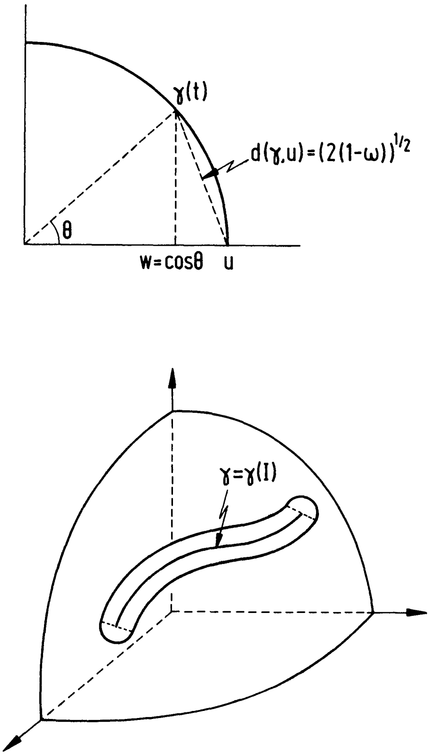

FIG. 1. *Tube of angular radius θ about curve γ, and a cross-sectional view.*

THEOREM 2.1 (Hotelling). *Let γ be a regular closed curve in S^{d-1} with length $|\gamma|$. If θ is sufficiently small,*

$$V(\gamma^{\theta}) = |\gamma|\Omega_{d-2}\sin^{d-2}\theta. \tag{2.1}$$

If γ is not closed, then the right-hand side of (2.1) *has an extra term,*

$$\omega_{d-2}\int_{\cos\theta}^{1}(1-z^2)^{(d-3)/2}\,dz. \tag{2.1'}$$

Here $\Omega_{d-2} = \pi^{(d-2)/2}/\Gamma(d/2)$ is the volume of the unit ball in $\mathbb{R}^{d-2}$, and $\omega_{d-2} = 2\pi^{(d-1)/2}/\Gamma((d-1)/2)$ is the $(d-2)$-dimensional volume of S^{d-2}.

Thus the volume for closed curves equals the length of the curve multiplied by the volume of a (cross-sectional) ball of dimension $d-2$ and radius $(1-w^2)^{1/2} = \sin\theta$. For nonclosed curves, the tube γ^{θ} includes hemispherical "caps" of dimension $d-1$, subtending an angle θ at each end. Hotelling notes that a necessary condition for the result to be exact is that there be no *local* self-overlap of the tube, i.e., $\sin\theta < \rho$, where ρ is the minimal radius of curvature of γ considered as a subset of $\mathbb{R}^d$. He also notes that for exactness there must be no self-intersections of the tube—an essentially *global* property. Section 4 discusses a numerical method for determining the largest radius for which no overlap occurs.

An analogous formula holds for the volume of tubes about closed curves α: $I \to \mathbb{R}^d$ in Euclidean space: If α^ρ denotes the tube of radius ρ about α, then in the absence of overlap $V(\alpha^\rho) = |\alpha|\Omega_{d-1}\rho^{d-1}$, with an extra term in $\Omega_d\rho^d$ for nonclosed curves. Although it is simpler to state and prove [e.g., Johnstone and Siegmund (1989)], this formula has fewer statistical applications than (2.1).

It is often convenient to express the volume $V(A)$ of a set $A \subset S^{d-1}$ in terms of probabilities for a random vector U uniformly distributed over the surface of S^{d-1}. Thus, $V(A) = V(S^{d-1})P(U \in A) = \omega_{d-1}P(U \in A)$. The projection of U on the closest unit vector in the curve γ leads to a new random variable,

$$W = \sup_{t \in I} \gamma(t)'U. \tag{2.2}$$

Since the event $\{W \geq w = \cos\theta\}$ is equivalent to $\{U \in \gamma^\theta\}$, results about tube volumes are equivalent to tail probability statements for W.

COROLLARY 2.2. *If γ is a nonclosed regular curve in S^{d-1}, then for w close to* 1,

$$P(W \geq w) = \frac{|\gamma|}{2\pi}(1 - w^2)^{(d-2)/2} + \frac{1}{2}P\left(B\left(\frac{1}{2}, \frac{d-1}{2}\right) \geq w^2\right) \tag{2.3}$$

where $B(1/2, (d-1)/2)$ is a random variable following the beta distribution. If γ is closed then the second term on the right-hand side is dropped.

Denote the "caps-adjusted" right-hand side of (2.3) by $b_\gamma(w)$. Using a method quite separate from that of Hotelling, Naiman (1986) has shown the remarkable result that this is *always* an upper bound.

THEOREM 2.3 (Naiman). *If γ is a piecewise regular curve of finite length in S^{d-1}, then for all $w \in [0, 1]$, $P(W \geq w) \leq b_\gamma(w)$.*

Note that the caps are needed for validity of the bound at all radii, *even if γ is a closed curve* (consider a large tube about a circle of small radius in S^2). Of course, this implies that $b_\gamma(w)$ cannot be sharp for closed γ and w near 1. We therefore reserve the term "Hotelling probability" $h_\gamma(w)$ for the expression that is exact for small radii,

$$h_\gamma(w) = \begin{cases} \dfrac{|\gamma|}{2\pi}(1 - w^2)^{(d-2)/2} & \text{if } \gamma \text{ is closed,} \\ b_\gamma(\omega) & \text{if not.} \end{cases} \tag{2.4}$$

If the curve has corners, then (2.4) is not exact, even for arbitrarily small radii. The error in (2.4) induced by a single discontinuity at $\dot{\gamma}(s)$ may be shown to be $O(\phi^3)$ as $\phi = \cos^{-1}(\dot{\gamma}(s-)'\dot{\gamma}(s+)) \to 0$, and hence will not be large for small kinks.

The analogue of Naiman's inequality for tubes about curves in Euclidean space had in fact been derived, as a technical ingredient of a larger calculation, by Estermann (1926). Hotelling's argument does not yield the Estermann–Naiman inequalities, nor do the methods of Estermann and Naiman allow one to obtain the exact volume of tubes of small radius. Johnstone and Siegmund (1989) present unified derivations of the results of Hotelling and Naiman via two distinct approaches.

B. *Three areas of application.* Theorem 2.1 applies directly to significance tests of a nonlinear parameter in regression, Hotelling's original setting. Here the sphere has dimension one less than the sample size. The second and third applications are to the related topics of simultaneous confidence and prediction bands. In these cases, the curves lie in a sphere of dimension one less than the number of parameters. Here the Hotelling–Naiman bound must be integrated against a χ^2 or F distribution over all tube radii. We recall these settings here: succeeding subsections connect to the tubes viewpoint and discuss the conservative bands that result.

(i) *Significance tests for a nonlinear parameter in regression.* Hotelling considered models of the form

$$Y_i = \alpha' z_i + \beta\lambda_i(\tau) + \epsilon_i, \qquad i = 1, \dots, n \tag{2.5}$$

where α, β, τ are fixed unknown parameters, $\lambda_i(\cdot)$ are known functions and ϵ_i are i.i.d. $N(0, \sigma^2)$. Examples for $\lambda_i(\tau)$ include $\lambda_i(\tau) = e^{\tau x_i}$ (Hotelling) and $\lambda_i(\tau) = x_i/(\tau + x_i)$ [Johansen (1984), Chapter 6]. The null hypothesis is absence of the nonlinear term, $\beta = 0$. Hotelling shows how to reduce (2.5) to a model in which the linear term $\alpha' z_i$ is absent, so we shall assume for simplicity that this has been done. Hotelling fits the parameters by least squares and arrives at the test statistic

$$L = \inf_\tau \frac{\Sigma_i (Y_i - \hat{\beta}_\tau \lambda_i(\tau))^2}{\Sigma_i Y_i^2}, \tag{2.6}$$

which is also the likelihood ratio test statistic for the indicated null hypothesis. Here $\hat{\beta}_\tau = Y'\lambda(\tau)/|\lambda(\tau)|^2$ is the regression of Y on $\lambda(\tau)$ for fixed τ, so (2.6) becomes

$$L = 1 - \sup_\tau \left(\frac{\lambda(\tau)'Y}{|\lambda(\tau)|\,|Y|} \right)^2 = 1 - \sup_\tau (\gamma(\tau)'U)^2.$$

Under the null hypothesis, $U = Y/|Y|$ is uniformly distributed on S^{n-1} and $\gamma(\tau) = \lambda(\tau)/|\lambda(\tau)|$ describes a curve in S^{n-1}. Thus, the likelihood ratio test, which rejects when $L \le \ell_0$, is equivalent to a test based on (2.2), namely,

$W \geq w_0$. Thus, the null hypothesis is rejected if $U = Y/|Y|$ falls in a sufficiently small tube about the curve $\gamma(\tau) = \lambda(\tau)/|\lambda(\tau)|$ (Figure 1). The P-value associated with such a tube can be calculated exactly for large w_0 and, furthermore, can always be bounded above by Theorem 2.3.

Keeping (1951) discussed the example $\lambda_i(\tau) = e^{\tau x_i}$, complete with bounds for nonoverlap of the tube. [Note, however, an apparent error at (7.2).] The detailed calculations could now, of course, be handled by numerical integration. Knowles and Siegmund (1989) treat $\lambda_i(\tau) = (x_i - \tau)_+$, which arises in change-point models in regression.

(ii) *Simultaneous confidence bands in regression.* Consider the Gaussian curvilinear regression model

$$(2.7) \qquad Y_i = \sum_{j=1}^{d} \beta_j a_j(t_i) + \epsilon_i, \qquad i = 1,\dots,n\ (> d),$$

where the components of $\beta = (\beta_1,\dots,\beta_d)$ are fixed unknown parameters, the components of $a(t) = (a_1(t),\dots,a_d(t))$ are known real valued functions, $\{t_i\}$ are fixed real numbers and $\{\epsilon_i\}$ are i.i.d. $N(0,\sigma^2)$ measurement errors. For simplicity, assume that σ^2 is known: Section 2E lists the minor changes when σ^2 must be estimated. Assume also that the $n \times d$ design matrix $A = (a_j(t_i))$ has full rank, so that the least squares estimate $\hat{\beta} \sim N_d(\beta, \sigma^2\Sigma)$, where $\Sigma = (A^tA)^{-1}$. A natural example is polynomial regression, for which $a_j(t) = t^{j-1}$ [e.g., Wynn and Bloomfield (1971)].

Many authors have constructed simultaneous confidence bands to take advantage of constraints on the predictor variables. Naiman (1986, 1990) lists some key references and gives a brief survey. In particular, Knafl, Sacks and Ylvisaker (1985) derive confidence bands via a discrete upcrossings method: Its relation to the tubes approach emerges from the upcrossings formulation of the latter in Johnstone and Siegmund (1989).

We focus on simultaneous confidence bands for the regression function $\beta' a(t)$ over a *fixed interval* $I \subset \mathbb{R}$. Scheffé-type bands have the form $\hat{\beta}' a(t) \pm c\sigma(a(t)'\Sigma a(t))^{1/2}$ for $t \in I$. The positive constant c is to be chosen to make the coverage probability

$$(2.8) \qquad P_{\beta,\sigma,\Sigma}\Big(\big|\beta' a(t) - \hat{\beta}' a(t)\big| \leq c\sigma\big(a(t)'\Sigma a(t)\big)^{1/2}\ \forall\, t \in I\Big)$$

close to some prespecified level, regardless of the value of β and σ^2. If $C = \{a(t)\colon t \in I\}$, then (2.8) becomes

$$(2.9) \qquad P_{\beta,\sigma,\Sigma}(T \leq c\sigma),$$

where

$$(2.10) \qquad T = T(\hat{\beta},\beta) = \sup_{a \in C} \frac{\big|a'(\hat{\beta} - \beta)\big|}{(a'\Sigma a)^{1/2}}.$$

We pursue this form further in 2C below.

(iii) *Prediction bands in random coefficient regression.* Consider a model of the form

$$Y_i = \sum_{j=1}^{d} B_j a_j(t_i) + \epsilon_i, \qquad i = 1, \dots, n\ (> d). \tag{2.11}$$

For example, $\{Y_i\}$ might be measurements of blood pressure on a single patient at n different times $\{t_i\}$. Let the measurement errors $\{\epsilon_i\}$ be i.i.d. $N(0, \sigma^2)$, the coefficients $B = (B_1, \dots, B_d)$ be *random* and distributed as $N_d(\beta, \Gamma)$, and let B be independent of $\{\epsilon_i\}$. As before, the components of $a(t) = (a_1(t), \dots, a_d(t))$ are known real functions, $\{t_i\}$ are fixed points in a particular interval $I \subset \mathbb{R}$ of interest, and the $n \times d$ matrix $A = (a_j(t_i))$ is taken to have full rank.

The vector of coefficients B characterize the individual, and their random variation describes the variation in the population (of patients in our example) of the individual mean functions. Since $Y = (Y_1, \dots, Y_n) \sim N_n(A\beta, A\Gamma A' + \sigma^2 I)$, its components are correlated even though the measurement errors $(\epsilon_i\}$ are independent.

The basic reference to such models is Rao (1965); for a recent survey see Spjøtvoll (1977), for a bibliography, Johnson (1977, 1980) and for a simple exposition, Johansen (1984).

Motivated by the gait analysis illustration in Section 3, we assume that many patients in the population have been investigated in the past and so the parameters $(\beta, \Gamma, \sigma^2)$ are considered known. Now a new patient arrives: Is her blood pressure versus time relation consistent with the "normal" population? One approach estimates the new patient's mean function $\hat{\beta}'a(t)$ using least squares estimates $\hat{\beta}$ from the new data $\{(t_i, Y_i)\}$ and model (2.11). If $B \sim N_d(\beta_1, \Gamma)$ for the new patient, then $\hat{\beta} \sim N_d(\beta_1, \Sigma)$ has two components of variance, $\Sigma = \Gamma + \sigma^2(A'A)^{-1}$. Plot her estimated mean function along with the prediction bands $\beta'a(t) \pm c(a(t)'\Sigma a(t))^{1/2}$ and flag the patient as abnormal if $\hat{\beta}'a(t)$ lies outside the prediction bands at *any* point $t \in I$. We choose c to control the chance of incorrectly flagging a normal patient and so need to calculate

$$P_{\beta, \sigma, \Gamma}\left[\hat{\beta}'a(t) \in \beta'a(t) \pm c\left(a(t)'\Sigma a(t)\right)^{1/2}\ \forall\, t \in I\right],$$

essentially (2.8) occurring in the confidence interval problem!

The prediction regions here and in Section 3 are admittedly ad hoc. However, the resulting random variable $T(\hat{\beta}, \beta)$ of (2.10) arises also as the likelihood ratio test statistic of the null hypothesis that $\beta_1 = \beta$ versus $\beta_1 = \beta + \rho\Sigma a$, for some $a \in C$ and $\rho \in \mathbb{R}$. Although this formulation yields a perhaps unnatural alternative, it does exhibit the kinds of departures from normalcy to which this approach is sensitive.

C. *Reducing the applications to Hotelling–Naiman.* To proceed with the last two examples, suppose that X is distributed as $N_d(\xi, \Sigma)$ with Σ known. Let $C \subset \mathbb{R}^d$ denote a set of vectors specifying linear combinations of interest to us: In our applications, C will usually be a curve. We want to make simultaneous confidence statements about $\{a'\xi, a \in C\}$ and to form prediction sets for the random variables $\{a'X, a \in C\}$.

In either case we start from the random variable

$$T = T(X, \xi) = \sup_{a \in C} \frac{a'(X - \xi)}{(a'\Sigma a)^{1/2}}. \tag{2.12}$$

If we can find the $P_{\xi,\Sigma}$ distribution of T, we can construct a $1 - \epsilon$ confidence set R_X,

$$R_X = \left[\{a'\xi\}_{a \in C} | T(X, \xi) \leq c_{1-\epsilon}\right],$$

where $c_{1-\epsilon}$ is the $1 - \epsilon$ quantile in the distribution of T. It is easily seen that the random set R_X covers the point $\{a'\xi\}_{a \in C}$ with $P_{\xi,\Sigma}$ probability $1 - \epsilon$.

Similarly a $1 - \epsilon$ prediction set,

$$R_\xi = \left[\{a'X\}_{a \in C} | T(X, \xi) \leq c_{1-\epsilon}\right],$$

contains the random point $\{a'X\}_{a \in C}$ with $P_{\xi,\Sigma}$ probability $(1 - \epsilon)$.

The variable T of (2.12) decomposes into

$$T = RW, \tag{2.13}$$

where $R^2 = (X - \xi)'\Sigma^{-1}(X - \xi)$ and

$$W = \sup_{a \in C} \frac{a'(X - \xi)}{(a'\Sigma a)^{1/2}((X - \xi)'\Sigma^{-1}(X - \xi))^{1/2}}$$

$$= \sup_{a \in C} \frac{(\Sigma^{1/2}a)'\Sigma^{-1/2}(X - \xi)}{|\Sigma^{1/2}a| \, |\Sigma^{-1/2}(X - \xi)|}.$$

If we define

$$\gamma(a) = \frac{\Sigma^{1/2}a}{|\Sigma^{1/2}a|}$$

and

$$U = \frac{\Sigma^{-1/2}(X - \xi)}{|\Sigma^{-1/2}(X - \xi)|},$$

then $\gamma = \gamma(C)$ is a subset of S^{d-1} and U is uniformly distributed on S^{d-1}. The distributions of R and W do not depend on ξ, Σ (except through γ), so for the remainder of this section we write simply P for $P_{\xi,\Sigma}$. The random variable R^2 is independent of W and follows a $\chi^2_{(d)}$ distribution. Thus,

$$P(T > c) = \int_c^\infty P(W > cr^{-1}) P(R \in dr). \tag{2.14}$$

In particular, when γ is a curve, W is a random variable of the form (2.2), to which the Hotelling–Naiman results apply.

Formula (2.14) says explicitly that as c increases, the tube in S^{d-1} gets narrower as the confidence/prediction bands become wider. Thus, narrow tubes do *not* correspond to narrow confidence bands.

D. *Upper bounds and their accuracy.* For the next two subsections, γ denotes a curve. Naiman employs his bound $b_\gamma(w)$ defined in (2.3) to bound (2.14) above by

$$(2.15) \qquad \int_c^\infty \min\{b_\gamma(cr^{-1}), 1\} P(R \in dr).$$

Knowles (1987) relaxes the constraint 1 in (2.15) and integrates (2.14) exactly, obtaining the less sharp but simpler bound

$$(2.16) \qquad P(T > c) \le \frac{|\gamma|}{2\pi} e^{-c^2/2} + 1 - \Phi(c),$$

where Φ is the standard Gaussian cumulative. We will employ (2.16) in Sections 3 and 6. All these upper bounds use the caps-adjusted version (2.3) *whether or not* the curve γ is closed. Knowles [(1987), page 33] shows that (2.16) may also be derived from the theory of upcrossings for Gaussian processes.

Analogous calculations are clearly possible for the two-sided tail probability $P(|T| > c)$. These are based on the inequality $P(|W| > w) \le 2P(W > w)$ and hence are conservative to the extent of overlap of the tubes γ^θ and $-\gamma^\theta$. A numerical method for calculating the radius of this first overlap appears in Section 4.

Let us turn to the accuracy of these bounds in the tails (as $c \to \infty$). We exploit the exactness of the Hotelling formula for small tubes. Suppose that it is known (from the methods of Section 4, for example) that the first self-overlap of the tube γ^θ occurs at $w_0 = \cos \theta_0$. For both closed and nonclosed curves, the bound

$$(2.17) \quad P(T > c) \le \int_c^{c/w_0} h_\gamma(cr^{-1}) P(R \in dr) + \int_{c/w_0}^\infty b_\gamma(cr^{-1}) P(R \in dr)$$

incurs an error $\int_{c/w_0}^\infty [b_\gamma(cr^{-1}) - P(W > cr^{-1})] P(R \in dr)$. This error is easily bounded by $(2\pi)^{-1}|\gamma| e^{-c_0^2/2} + P(Z_1 \ge c, |Z| \ge c_0)$, where $Z = (Z_1, \dots, Z_d) \sim N_d(0, I)$ and $c_0 = c/w_0$. Thus, for a closed curve,

$$(2.18) \qquad P(T > c) = \frac{|\gamma|}{2\pi} e^{-c^2/2} + O\left(\left(\frac{\cdot c}{w_0} \right)^{d-2} e^{-c^2/2w_0^2} \right).$$

E. *Unknown* σ^2. Let $X \sim N(\xi, \sigma^2\Sigma)$ with Σ known and let V^2 be an independent estimate of σ^2 satisfying $\nu V^2/\sigma^2 \sim \chi^2_{(\nu)}$. In the definition of C_X and R_ξ, replace T by T/V and the distribution $P_{\xi,\Sigma}$ by P_{ξ,Σ,σ^2}. In the decomposition $T = RW$, now $R^2 = (X - \xi)'\Sigma^{-1}(X - \xi)/V^2$ and R^2/d follows an F distribution on (d, ν) degrees of freedom independently of W. With this change in (2.14) (the case actually considered by Naiman), (2.16) becomes

$$P(T > c) \leq \frac{|\gamma|}{2\pi}\left(1 + \frac{c^2}{\nu}\right)^{-\nu/2} + P(t_{(\nu)} \geq c),$$

where $t_{(\nu)}$ denotes a t-variate on ν degrees of freedom. Formulas (2.17) and (2.18) have corresponding extensions.

In these simultaneous confidence and prediction problems, a conservative procedure can always be obtained by Sheffé's method: replace the set $\gamma = \gamma(C)$ by all of S^{d-1}. We would thus expect useful reductions in width of the bands —smaller choices of $c_{1-\epsilon}$—when γ is a sparse subset of S^{d-1}, for example, if $|\gamma|$ is small or d is large, as occurs in polynomial or trigonometric regression. A crude upper bound for the reduction in $c_{1-\epsilon}$ comes by comparing the Scheffé value $c_{1-\epsilon} = [dF_{d,\nu}(1-\epsilon)]^{1/2}$ to the t statistic value $c_{1-\epsilon} = t_{(\nu)}(1-\epsilon)$ that would apply if γ reduced to a pair of antipodal points.

F. *A special case of Weyl's result.* Hotelling proposed the application of tube methods to extensions of (2.5) containing two or more nonlinear parameters. The associated volume problem for small tubes about manifolds without boundary was then solved by Weyl (1939) in a paper of considerable influence in differential geometry.

Briefly, Weyl's result is as follows. Let C be a κ-dimensional manifold contained in S^{d-1}. Let $\mu = d - 1 - \kappa$ denote the codimension of C in S^{d-1}. As in Section 2, the tube of geodesic radius θ about C in S^{d-1} is defined by $C^\theta = \{u \in S^{d-1}\colon \sup_{\gamma \in C} u'\gamma \geq \cos\theta\} = \{u \in S^{d-1}\colon d(u, C) \leq (2(1-w))^{1/2}\}$, where $w = \cos\theta$.

THEOREM 2.4 (Weyl). *If C is a smooth κ-dimensional manifold (without boundary) embedded in S^{d-1}, then for sufficiently small θ,*

$$V(C^\theta) = \omega_{\mu-1} \sum_{\substack{0 \leq e \leq \kappa \\ e \text{ even}}} k_e J_e(\theta), \qquad \mu = d - 1 - \kappa, \tag{2.19}$$

where $J_e(\theta)$ is defined by

$$\mu(\mu+2)\cdots(\mu+e-2)J_e(\theta) = \int_0^\theta (\sin\rho)^{\mu+e-1}(\cos\rho)^{\kappa-e}\, d\rho \tag{2.20}$$

and k_e are certain integral invariants of C, with k_0 in particular being the κ-dimensional surface area of C. [If $e = 0$, the coefficient of $J_e(\theta)$ is 1.]

Hotelling's formula for closed curves is the case $\kappa = 1$. Knowles and Siegmund (1989) present an account of Weyl's theorem tailored to statistical application, extensions to (two-dimensional) manifolds with boundary and detailed numerical examples. Naiman (1990) extends Weyl's theorem to spherical polyhedra of the kind occurring in multiple regression. Weyl's formula is related to, but distinct from, the Steiner formula for the volume of a parallel translate of a convex body [Santaló (1976)].

For sufficiently small angles θ, the leading term ($e = 0$) dominates in Weyl's formula (2.19). This approximation involves only the surface area of the manifold, and we express it in a form convenient to our applications to projection pursuit regression in Section 6. As in (2.3), division by the volume of S^{d-1} converts (2.19) into a probability statement about a uniform random vector U on S^{d-1}. If $W = \sup_{\gamma \in C} \gamma' U$, then

$$P(W \geq \cos\theta) \sim \frac{\omega_{\mu-1}}{\omega_{d-1}} \frac{k_0}{2} \int_{\cos^2\theta}^{1} (1-u)^{\mu/2-1} u^{(\kappa+1)/2-1}\, du \quad \text{as } \theta \to 0,$$

where we have set $u = \cos^2 \rho$. The integral is proportional to the tail of a Beta($(\kappa+1)/2, \mu/2$) distribution. Since $U_1^2 + \cdots + U_{\kappa+1}^2 \sim$ Beta($(\kappa+1)/2$, $\mu/2$), we obtain the following, after collecting constants.

Corollary 2.5. *Under the conditions of Theorem* 2.4,

$$P(W \geq \cos\theta) \sim \frac{k_0}{\omega_\kappa} P\big(U_1^2 + \cdots + U_{\kappa+1}^2 \geq \cos^2\theta\big) \quad \textit{as } \theta \to 0.$$

3. An illustration in gait analysis. Olshen, Biden, Wyatt and Sutherland (1989) have used bootstrap methods for simultaneous prediction regions as part of an extensive study of normal and abnormal gait in children. The purpose of this section is to show, using some of Olshen et al.'s data, how Hotelling's approach approximately reproduces and illuminates the bootstrap analysis in three particular cases.

To study the walking cycle of a child, markers are placed on the pelvis and lower limbs to identify bony landmarks. For example, markers at the hip, knee and ankle define knee flexions. The child walks, and the motion is recorded using cine film or video. Measurements of various "joint rotations" are measured over one or more cycles. One goal is to characterize walking patterns in normal children and to develop diagnostic measures of abnormal gait. Further details of the data collection and modeling are given by Olshen et al. Measurements of, say, knee flexion for a particular child are assumed to follow the model

$$(3.1) \quad Y_i = A_0 + \sum_{j=1}^{p} A_j \cos(jt_i) + B_j \sin(jt_i) + \epsilon_i, \qquad i = 1, \dots, k.$$

The points $t_i = 2\pi(i-1)/k$, $i = 1, \dots,$ k divide the step cycle into k equal parts. In Olshen et al., k depends on the rotation being measured, but is at least 16. The measurement errors are assumed to be i.i.d. with mean zero and variance σ^2. As in Section 2B(iii), the vector of coefficients $\Xi = (A_0, A_1, \dots, A_p, B_1, \dots, B_p)$ is regarded as characteristic of a particular child and walk, and hence random, with mean $\xi = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p)$ and covariance matrix Γ, which is usually not diagonal. The errors $\{\epsilon_i\}$ and coefficients Ξ are assumed independent. Our theory further requires that $\{\epsilon_i\}$ and Ξ follow Gaussian distributions—Olshen et al.'s use of bootstrap methods deliberately avoids this assumption, placing in its stead greater reliance on the data at hand.

Suppose now that (ξ, Γ, σ^2) are known for a population of "normal" children: In practice these are estimated from a learning sample. A new child, with his own unknown value of Ξ in model (3.1), produces data $\{(t_i, Y_i)\}$. To decide if this child is normal, Olshen et al. compute least squares estimates $\hat{\Xi} = (\hat{A}_0, \dots, \hat{A}_p, \hat{B}_1, \dots, \hat{B}_p)$ of Ξ (with $p = 6$). In practice the variability of $\hat{A}_0$ is so large that it is found helpful to separate it from the remaining harmonic coefficients. Henceforth, we consider only $X = (\hat{A}_1, \dots, \hat{A}_p, \hat{B}_1, \dots, \hat{B}_p)$. Let $a(t) = (\cos t, \dots, \cos pt, \sin t, \dots, \sin pt)$. The child is flagged as abnormal if the estimated curve $a(t)'X$ lies *at any point* $t \in [0, 2\pi]$ outside the prediction bands

$$a(t)'\xi \pm cv^{1/2}(t). \tag{3.2}$$

Here $v(t)$ is the variability that would be expected for a normal child, $v(t) = \operatorname{Var}(a(t)'X|\xi, \Gamma, \sigma^2) = a(t)'[\Gamma_h + 2\sigma^2 k^{-1} I]a(t)$, where Γ_h is the submatrix of Γ corresponding to the harmonic components. The scaling factor c is chosen to yield a prescribed probability (such as 0.05) that a normal child [with coefficients chosen according to (ξ, Γ, σ^2)] will be incorrectly flagged as abnormal. Hence, we need to evaluate probabilities such as

$$P_{\xi, \Gamma, \sigma^2}\left\{|a(t)'(X - \xi)| \le c\left(a(t)'\Sigma a(t)\right)^{1/2} \text{ for all } t \in [0, 2\pi]\right\}, \tag{3.3}$$

where $\Sigma = \Gamma_h + 2\sigma^2 k^{-1} I$. This is precisely an expression of the form (2.9) and (2.10) for $I = [0, 2\pi]$ and can be studied as described after (2.12).

Olshen et al. compute $c_{0.05}$ via the bootstrap approximation to the $P_{\xi, \Gamma, \sigma^2}$ distribution of $T_{\text{ts}}(X, \xi) = \sup_{a \in a[0, 2\pi]} |a'(X - \xi)|/(a'\Sigma a)^{1/2}$ based on a learning sample of 39 normal children. (The subscript ts denotes two-sided.) We invoke the Gaussian assumptions and the decomposition $T = RW$ of (2.13) to study $c_{0.05}$ theoretically. We have $R^2 \sim \chi^2_{(12)}$ and, independently of R, $W = \sup_{t \in [0, 2\pi]} |\gamma(t)'U|$, where $\gamma(t) = \Sigma^{1/2} a(t)/|\Sigma^{1/2} a(t)|$ and U is uniformly distributed on $S^{11} \subset \mathbb{R}^{12}$.

The two-sided version of the Hotelling–Naiman–Knowles bound (2.16) is

$$P(T_{\text{ts}} \le c) \ge 1 - 2\left\{\frac{|\gamma|}{2\pi} e^{-c^2/2} + 1 - \Phi(c)\right\}. \tag{3.4}$$

The right-hand side of (3.4), which we shall denote $k_{\text{ts}}(c)$, is readily computable. As an example, we use the data (i.e., mean and covariance matrix) on left ankle dorsi-plantar flexion for 39 normal 5-year-olds obtained by Olshen et al. This measurement was chosen for ease of comparison with the plots presented by Olshen et al. (cf. their Figure 7). The covariance matrix Σ is far from diagonal: Consequently, the norm of $\gamma'(t)$ oscillates between 1.44 and 4.28 for t in $[0, 2\pi]$; crude numerical integration yields a total length $|\gamma|/2\pi = 2.496$.

The second column of Table 1 presents some representative values of $k_{\text{ts}}(c)$ corresponding to the right tail of the distribution of T_{ts}. By linear interpolation we arrive at a value of $c_{\text{tube}} = 3.05$ as an approximation to the 95th

TABLE 1

Simulated and approximate distribution of $T_{ts} = RW_{ts}$ for the 5-year-old left ankle dorsi-plantar flexion example. $k_{ts}(c)$ denotes the Knowles lower bound (3.4), $\hat{F}(c)$ the estimated distribution function of T_{ts} from 500 replications, as described in the text, $2\hat{\sigma}(c)$ the corresponding approximate half-width of the symmetric two-sided confidence interval for $F(c)$.

c	$k_{ts}(c)$	$\hat{F}(c)$	$2\hat{\sigma}(c)$	c	$k_{ts}(c)$	$\hat{F}(c)$	$2\hat{\sigma}(c)$
2.6	0.821	0.861	0.015	3.4	0.984	0.986	0.003
2.7	0.863	0.891	0.013	3.5	0.989	0.990	0.003
2.8	0.896	0.915	0.011	3.6	0.992	0.993	0.002
2.9	0.922	0.935	0.009	3.7	0.994	0.995	0.002
3.0	0.942	0.951	0.008	3.8	0.996	0.997	0.001
3.1	0.957	0.963	0.006	3.9	0.997	0.998	0.001
3.2	0.969	0.973	0.005	4.0	0.998	0.998	0.001
3.3	0.977	0.980	0.004				

percentile of the distribution of T_{ts}. Figure 2 plots the resulting normal-theory prediction regions together with the corresponding bootstrap prediction regions obtained by Olshen et al. (the bootstrap 95th percentile, $c_{boot} = 3.22$). The apparent close agreement suggests that in at least this setting Hotelling's theory provides a reasonable approximation. (Of course, the *shapes* of the prediction regions are prescribed by the mean and covariance matrix—it is the *widths* of the regions that concern us here.)

(Here the bootstrap percentile is obtained by interpolation from the quadratic regression fitted to the bootstrap frequencies of 0.89, 0.93, 0.97, 0.99 at $c = 2.75, 3.00, 3.5, 4.0$ provided by E. Biden. Since the bootstrap frequencies are themselves Monte Carlo estimates based on a relatively small number of replications, we do not attempt any refined analysis of the quality of approximation of c_{boot} and c_{tube}.)

The computations leading to Figure 2 were repeated for two other measurements on the same 39 normal 5-year-olds: left knee flexion and left hip flexion/extension. In terms of agreement between bootstrap and tube 95th percentiles, the knee flexion is closest ($c_{tube} = 3.05 < 3.06 = c_{boot}$), and the hip flexion least close ($c_{tube} = 2.95 < 3.15 = c_{boot}$). Of course, the tube method as applied here does not make allowance for the variability in the estimates of (ξ, Γ, σ^2) that Olshen et al. compute from the learning sample of size $39k$. To do so would presumably lead to slightly wider bands, although in the prediction setting this is a lower-order effect (asymptotically in the size of the learning sample, here at least $16 \times 39 = 624$).

What happens at other percentiles? Moving in from the tails, the bound $k_{ts}(c)$ will become more conservative due to self-overlap and mutual overlap of the tubes γ^{θ} and $-\gamma^{\theta}$. We want to compare the bootstrap distribution to the normal theory benchmark $F(c) = P(T_{ts} \leq c)$. Of course, the exact value of $F(c)$ is unknown, but can be efficiently simulated because of the decomposition $T_{ts} = RW_{ts}$. Indeed, if $W_1, \dots, W_N$ are i.i.d. observations on W_{ts}, and $\tilde{G}(y) = P(\chi^2_{(12)} > y)$, then an unbiased estimator of $\hat{F}(c)$ is given by $1 -$

$N^{-1}\Sigma_1^N\tilde{G}(c^2W_i^{-2})$. Columns 3 and 4 of Table 1 present unbiased estimates and half-widths of approximate 95% two-sided symmetric confidence intervals for $P(T_{\text{ts}} \le c)$ based on $N = 500$ simulations. For $c > 3.0$ (corresponding to tail probabilities of 0.05 or smaller), the lower bound $k_{\text{ts}}(c)$ lies within the confidence interval. In interpreting Table 1, it is important to note [from the definition of $\hat{F}(c)$] that the rows are positively correlated, perhaps strongly so. Note, of course, that these results depend on the particular curve γ (and in this example on the covariance matrix for the left ankle measurements).

We conclude with a purely qualitative observation. Figure 3 shows the bootstrap distribution of T_{ts} obtained by Olshen et al. together with the simulated normal-theory distribution of T_{ts} obtained as described above. Although most of the plot lies out of the intended range of application of approximation (3.4), it does show that the sigmoidal shape noted by Olshen et al. corresponds to the sigmoidal shape of the $\chi_{(12)}$ distribution function scaled by the random multiplier W_{ts} which in this case had 95% of its mass concentrated in the range 0.37 to 0.81.

4. When do tubes overlap?

We have seen that Hotelling's formula provides an expression for the volume of a tube that is exact for sufficiently small tube radii and an upper bound for all radii. Further, a bound for the error in Hotelling's formula when averages are taken over all radii can be derived from knowledge of the critical radius ρ_c at which self-intersection occurs. Since ρ_c depends on global features of the curve, it must generally be found numerically. This section shows how this computation is reduced to a relatively simple bivariate (sometimes univariate) optimization. Examples are given in Section 5. Although we are chiefly interested in curves embedded in spheres, we begin with the Euclidean case for simplicity.

Curves in $\mathbb{R}^d$. Suppose that α: $[0, c] \to \mathbb{R}^d$ is a regular *closed* curve. The tube of radius ρ about α is denoted by $\alpha^\rho = \{x \in \mathbb{R}^d\colon \min_t |x - \alpha_t| \le \rho\}$. The *cross-section* of the tube α^ρ at α_t is defined by

$$C_\rho(\alpha_t) = \{x \in \alpha^\rho\colon x - \alpha_t \perp \dot{\alpha}_t\}$$

(dots denote differentiation with respect to t). We often omit the subscript ρ.

A point $x \in \alpha^\rho$ lies in at least one cross-section: At a minimum of the function $t \to |x - \alpha_t|^2$, one has $(x - \alpha_t)' \cdot \dot{\alpha}_t = 0$. Thus

$$\alpha^\rho = \bigcup_{t \in [0, c]} C(\alpha_t). \tag{4.1}$$

We shall say that *no self-overlap* of the tube occurs if the union in (4.1) is *disjoint*. The *critical radius* ρ_c of first overlap is defined as

$$\rho_c = \inf\{\rho \ge 0\text{: self-overlap occurs}\}.$$

(The infimum need not be attained, consider an ellipse in the plane.) It is shown in Johnstone and Siegmund (1989), that the Hotelling formula $V(\alpha^{\rho}) = |\alpha|\Omega_{d-1}\rho^{d-1}$ is exact if $\rho \le \rho_c$.

ρ_c may be expressed as the minimum of a bivariate function.

PROPOSITION 4.1. *If α: $[0, c] \to \mathbb{R}^d$ is regular and closed,*

$$\rho_c = \inf_{s,t} \frac{|\alpha_s - \alpha_t|^2}{2|P_{\dot{\alpha}_t^{\perp}}(\alpha_s - \alpha_t)|} \triangleq \inf_{s,t} \frac{d^2}{2L}, \tag{4.2}$$

where $P_{\dot{\alpha}_t^{\perp}}$ denotes orthogonal projection on the hyperplane normal to $\dot{\alpha}_t$.

PROOF. If $\rho > \rho_c$, there exist s, t and $x \in C(\alpha_s) \cap C(\alpha_t)$. By shrinking ρ and relabeling if necessary, we may assume that $|x - \alpha_t| = \rho$ and that β, the projection of α_s onto the hyperplane through α_t normal to $\dot{\alpha}_t$, is nonzero (see Figure 4). Now alter x such that $x - \alpha_t$ is collinear with $\beta - \alpha_t$, giving $x = \alpha_t + \rho(\beta - \alpha_t)/|\beta - \alpha_t|$: As this only moves x closer to α_s, we still have $|x - \alpha_s| \le \rho$. We use abbreviations $d = |\alpha_s - \alpha_t|$, $L = |\beta| = |P_{\dot{\alpha}_t^{\perp}}(\alpha_s - \alpha_t)|$, $\rho = |x - \alpha_t|$ and $V = |\alpha_s - \beta|$. Since overlap occurs, $|x - \alpha_s|^2 = V^2 + (\rho - L)^2 \le \rho^2$. Since $V^2 = d^2 - L^2$, this is algebraically equivalent to $d^2 \le 2L\rho$. Thus, if $\rho > \rho_c$ then $\rho \ge d^2/2L$, which shows that $\rho_c \ge I$, the infimum occurring in Proposition 4.1. Conversely, if $I < \rho_c$, there would be points α_s, α_t with $d^2/2L = \rho' < \rho_c$. We would construct x as in Figure 4 with $|x - \alpha_t| = \rho'$ and by reversing the algebra above, $|x - \alpha_s| = \rho'$. Since overlap cannot occur at radius $\rho' < \rho_c$, it follows that $I = \rho_c$, which completes the proof. □

Some feeling for the content of (4.2) can be obtained from the special case in which the infimum is attained at two points s, t such that $\alpha_s - \alpha_t$ is perpendicular to both $\dot{\alpha}_s$ and $\dot{\alpha}_t$. It is now natural to take $x = (\alpha_s + \alpha_t)/2$, since $d = L$ in Figure 4.1, $d^2/2L$ reduces to the value we would intuitively expect, namely $d/2$. This is an example of *nonlocal* overlap and should be contrasted with the *local* overlap case in which the tube radius ρ exceeds the radius of curvature $\rho(s)$ of α at point $\alpha(s)$. Formula (4.2) must include this situation also, and indeed one may check that

$$\lim_{t \to s} \frac{d^2}{2L} = \rho(s). \tag{4.3}$$

These considerations suggest the following more perspicuous formula for ρ_c.

PROPOSITION 4.2. *Suppose that α: $[0, c] \to R^d$ is regular, closed and C^2. Let $E = \{(s, t)$: $s \ne t$ and (s, t) is a critical point of the function $(s, t) \to |\alpha_s - \alpha_t|^2\}$. Then*

$$\rho_c = \min\left\{\inf_s \rho(s), \inf_E |\alpha_s - \alpha_t|/2\right\}.$$

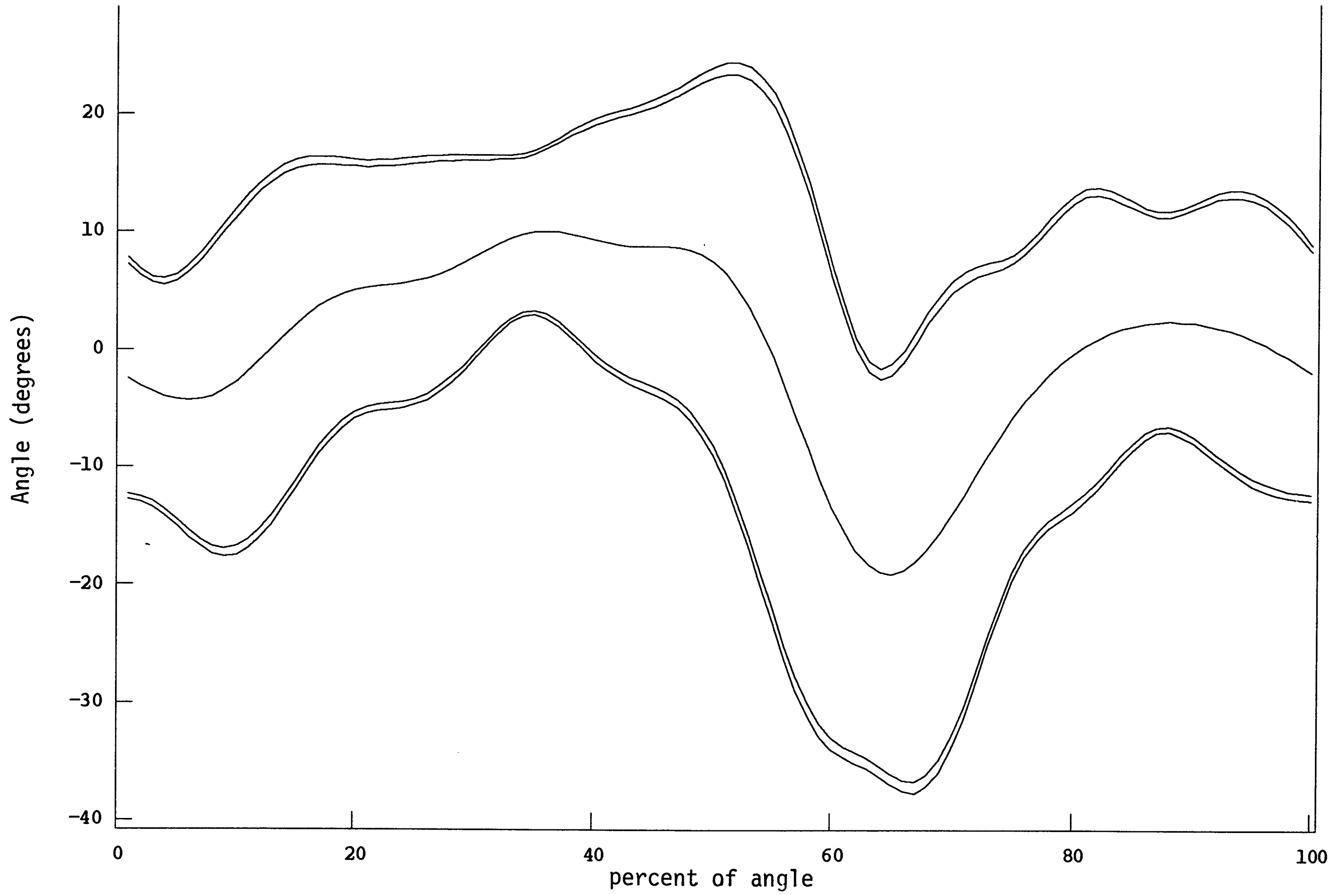

FIG. 2. *95% prediction regions for 5-year-old left ankle dorsi-plantar flexion about a central mean function estimated from a learning sample of* 39 *in Olshen, Biden, Wyatt and Sutherland* (1989) *of walking cycle, ordinate is angle of rotation* (*in degrees*). *Interior prediction regions are the tube approximation described in Section* 3, *the slightly more conservative exterior regions are the bootstrap curves produced by Olshen, Biden, Wyatt and Sutherland* ($c_{\text{tubes},0.95} = 3.05$, $c_{\text{boot},0.95} =$

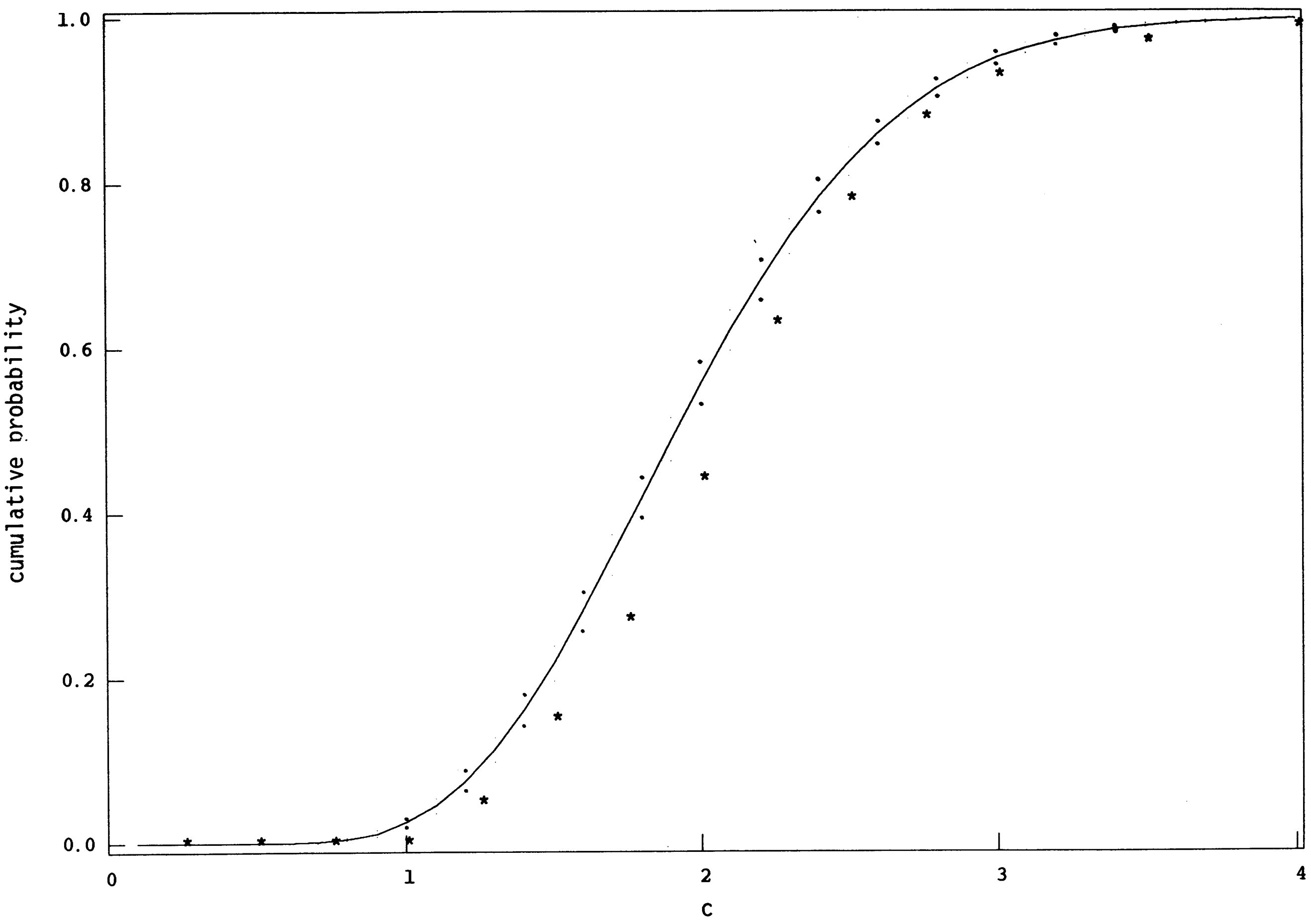

FIG. 3. *Solid line*: *simulated* (500 *replications*) *distribution of* $T_{ts} = RW_{ts}$ *for* 5-*year-old ankle dorsi-plantar flexion from normal theory assumptions, with some representative* 95% *confidence intervals*. *Asterisks*: *bootstrap distribution of T obtained by Olshen*, *Biden*, *Wyatt and Sutherland* (1989) (*uncertainties here are not shown*). *Abscissa is cutoff value c*, *ordinate is cumulative probability*.

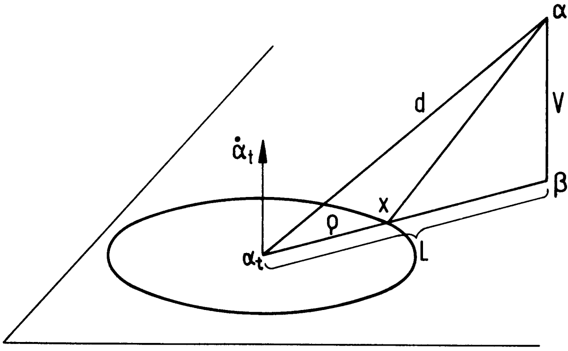

FIG. 4. *Schematic for proof of Proposition* 4.1.

This version exhibits local and global components quite explicitly. However, it is doubtful whether this formula is easier to implement computationally than (4.2).

PROOF OF PROPOSITION 4.2. Property (4.3) shows that the right side is an upper bound to ρ_c. If $\rho_c = \rho_*(:= \inf_s \rho(s)$, then equality is immediate, so it remains to consider the case in which $\rho_c < \rho_*$. In this case, any global minimum (s_0, t_0) of the lower semicontinuous function $\varphi(s,t) = \frac{1}{2}|\alpha_s - \alpha_t|^2/|P_{\dot{\alpha}_t^\perp}(\alpha_s - \alpha_t)|$ has $s_0 \neq t_0$, and it suffices to show that one such minimum $(s_0, t_0) \in E$.

We choose $\rho_n \downarrow \rho_c$ and (s_n, t_n, x_n) such that $x_n \in C_{\rho_n}(\alpha_{s_n}) \cap C_{\rho_n}(\alpha_{t_n})$ and (using the method of Proposition 4.1) $\varphi(s_n, t_n) \le \rho_n$. By compactness and lower semicontinuity of φ, a limiting triple (s, t, x) exists for which $\varphi(s,t) = \rho_c$ and $s \neq t$. Further $(x - \alpha_s)'\dot{\alpha}_s = 0$ and $|x - \alpha_s| \le \rho_c < \rho(s)$, which shows that s is a *strict* local minimum of the function $u \to |x - \alpha_u|^2$. A similar conclusion holds for t, so there exist local *maxima* u_i $(i = 1,2)$ between s and t for which $|x - \alpha_{u_i}| \ge \rho_* > \rho_c$.

As in the proof of Proposition 4.1, we may assume that $x - \alpha_t$ lies along $P_{\dot{\alpha}_t^\perp}(\alpha_s - \alpha_t)$. We argue by contradiction that $(\alpha_s - \alpha_t)'\dot{\alpha}_t = 0$. If not, consider the triangle with vertices α_t, α_s, and x. Move x along the perpendicular to $\alpha_s - \alpha_t$ to a point $\tilde{x}$ with $|\tilde{x} - \alpha(s)| < \rho_c$ and $|\tilde{x} - \alpha(t)| < \rho_c$ but with $|\tilde{x} - x|$ small enough that $|\tilde{x} - \alpha(u_i)| > \rho_c$. Hence, there must be two *distinct* local minima $\tilde{s}, \tilde{t}$ of $u \to |\tilde{x} - \alpha(u)|^2$ with values less than ρ_c^2. Thus, $\tilde{x} \in C_\rho(\alpha(\tilde{s})) \cap C_\rho(\alpha(\tilde{t}))$ for some $\rho < \rho_c$, which contradicts the definition of ρ_c. By reversing the roles of s and t, we also have $(\alpha_s - \alpha_t)'\dot{\alpha}_s = 0$, which completes the proof. □

The extension of Proposition 4.1 to cover smooth nonclosed curves is straightforward. The argument there works for $t \in (0, c)$ and $s \in [0, c]$, and so symmetry considerations dictate that it is only necessary to check $|\alpha_0 - \alpha_c|$ also.

Proposition 4.1′. *If α: $[0, c] \to \mathbb{R}^d$ is smooth but not closed, then*

$$\rho_c = \min\left\{ \inf_{s,t\in[0,c]^2} \frac{d^2}{2L}, \frac{|\alpha_0 - \alpha_c|}{2} \right\}.$$

Curves in spheres. We turn now to smooth closed curves γ: $[0, c] \to S^{d-1}$. The tube of radius $w = \cos\theta$ is denoted by

$$\gamma^\theta = \left\{ y \in S^{d-1}: \sup_t y'\gamma_t \geq \cos\theta \right\} = \left\{ y \in S^{d-1}: \inf_t |y - \gamma_t| \leq [2(1-w)]^{1/2} \right\}.$$

(Recall Figure 1). The cross-section of γ^θ at γ_t is defined by

$$C(\gamma_t) = \{ y \in S^{d-1}: y = \gamma_t \cos\phi + v \sin\phi,$$
$$\text{where } 0 \leq \phi \leq \theta \text{ and } |v| = 1, v \perp \gamma_t, v \perp \dot{\gamma}_t \}.$$

As in the Euclidean case, $\gamma^\theta = \bigcup_{t\in[0,c]} C(\gamma_t)$, and we say that self-overlap occurs if the union is not disjoint. The *critical angle* of the first overlap is $\theta_c = \inf\{\theta \geq 0$: self-overlap occurs$\}$, and the spherical version of Hotelling's formula is exact iff $\theta \leq \theta_c$.

To give an expression for all θ_c, let P_{M_t} denote projection onto the subspace M_t spanned by γ_t and $\dot{\gamma}_t$.

Proposition 4.3. *If γ: $[0, c] \to S^{d-1}$ is regular and closed,*

$$\cot^2\theta_c = \sup_{s,t} \frac{1 - \gamma_s' P_{M_t} \gamma_s}{(1 - \gamma_s'\gamma_t)^2} \triangleq \sup_{s,t} h(s,t). \tag{4.4}$$

If γ is not closed, θ_c must be reduced to $2^{-1}\cos^{-1}(\gamma_0'\gamma_c)$ if the latter is smaller.

Proof. The argument is similar to that for Proposition 4.1. If $\theta \geq \theta_c$, then pick s, t and $y \in C(\gamma_s) \cap C(\gamma_t)$ such that $y = \gamma_t \cos\theta + v\cos\theta$, with $|v| = 1$ and $v \in M_t^\perp$. There is no loss in assuming also that v is chosen to minimize $|y - \gamma_s|^2$, which forces $v_{\text{opt}} = P_{M_t^\perp}\gamma_s / |P_{M_t^\perp}\gamma_s|$. By assumption $|y - \gamma_s|^2 \leq 2(1 - \cos\theta)$, and some algebra reveals that this inequality is equivalent to

$$\cos\theta(1 - \gamma_s'\gamma_t) \leq \sin\theta |P_{M_t^\perp}\gamma_s|.$$

The argument is completed in the same fashion as for Proposition 4.1. □

A case of nonlocal overlap occurs when the supremum in (4.4) is attained at two points s, t with $\gamma_s - \gamma_t$ perpendicular to $\dot{\gamma}_s$ and $\dot{\gamma}_t$. In this case one expects [and can verify from (4.4)] that $\gamma_s'\gamma_t = \cos 2\theta_c$. To see the connection with the local quantities, it is entertaining to compute that

$$\lim_{t\to s} h(s,t) = k_g^2(s),$$

where $k_g^2(s)$ is the squared geodesic curvature of γ at s, given explicitly by $k_g^2(s) = |(I - P_{M_s})\ddot{\gamma}_s|^2 / |\dot{\gamma}_s|^4$. Since both numerator and denominator of (4.4) are $O((t-s)^4)$ as $t \to s$, it is useful in numerical work to use $k_g^2(t)$ approximating $h(s,t)$ for t near s.

REMARK. Although our main interest in S^{d-1} is in angles rather than radii, we may rewrite (4.4) in a form similar to (4.2), since

$$h^{-1/2}(s,t) = \frac{|\gamma_s - \gamma_t|^2}{2\left|P_{M_t^{\perp}}(\gamma_s - \gamma_t)\right|}.$$

Shift-invariant case. Formula (4.4) simplifies to a univariate minimization in the special case when $\gamma_s'\gamma_t = g(s-t)$. Indeed, $\dot{\gamma}_s'\gamma_t = g'(s-t)$ and $|\dot{\gamma}_t|^2 = -g''(0) = r^{-1}$, say. Then $h(s,t) = k(s-t)$, where

$$k(u) = \frac{1 - g^2(u) - r(g'(u))^2}{[1-g(u)]^2}, \tag{4.5}$$

and

$$k(0) = k_g^2 = g_2^{-2} g_4 - 1, \tag{4.6}$$

where $g_i = g^{(i)}(0)$ and, in particular, $g_3 = 0$.

Antipodal tubes. In two-sided problems, we also wish to know when γ^θ first intersects its reflection $-\gamma^\theta$. A slight change to the proof of (4.4) shows that this angle θ_b arises as the solution to

$$\cot^2\theta_b = \sup_{s,t} \frac{1 - \gamma_s' P_{M_t}\gamma_s}{(1 + \gamma_s'\gamma_t)^2} \triangleq \sup_{s,t} h_b(s,t), \tag{4.7}$$

that is, by changing the sign of $\gamma_s'\gamma_t$ in (4.4) and by changing $[1-g(u)]^2$ to $[1+g(u)]^2$ in the denominator of (4.5) in the shift-invariant case.

REMARK (Projections of points and uniqueness of nonlinear least squares estimates). Define the projection $\pi(x)$ of a point x in $\mathbb{R}^d$ onto a curve α in $\mathbb{R}^d$ as any point $\alpha(t)$ minimizing $t \to |x - \alpha(t)|^2$. Propositions 4.1 and 4.1′ imply that if $|\pi(x) - x| < \rho_c$, then $\pi(x)$ is uniquely defined. Similarly, for $y \in S^{d-1}$ and a curve γ in S^{d-1}, the projection $\pi(y)$ is a value $\gamma(t)$ that minimizes geodesic distance to γ or, equivalently, maximizes $\sup_t \gamma(t)'y$. Proposition 4.3 gives a sufficient condition for the projection to be unique. These results lead to sufficient conditions for uniqueness of least squares estimates in the model (2.5) of Section 2. Again for simplicitly, assume that the $\alpha'z$ term is absent. If β is known, then the least squares estimate $\hat{\tau}$ is unique if $|Y - \beta\lambda(\hat{\tau})|^2 < \rho_c^2$, where we consider the curve α: $\tau \to \beta\lambda(\tau)$. If β is unknown, then by arguing as in Section 2A(ii), we find that the least squares estimator $(\hat{\beta}, \hat{\tau})$ is unique if $(\gamma(\hat{\tau})'U)^2 \geq \cos^2\theta_b$, where θ_b [defined in (4.7)] is determined from the curve $\gamma(\tau) = \lambda(\tau)/|\lambda(\tau)|$ and $U = Y/|Y|$. Some practical examples involving kinetic data to which these considerations are relevant appear in Johansen [(1984), Chapter 6]. Projections of points on curves and their uniqueness are also basic to the study of principal curves, a nonlinear extension of principal components [Hastie and Stuetzle (1989)].

5. Tours through multivariate data: Tubes illustrated on Andrews' plots. A natural way to view high-dimensional data is through low-dimensional projections, and when these projections are changed continuously, a curve in the space of projections results. If we allow that nearby projections yield similar views, then Hotelling's tube method offers a way to assess the fraction of all possible views that may be scanned using a particular curve.

An early example is the plot proposed by Andrews (1972), in which each data point $x = (x_1, \dots, x_d)$ is mapped into a trigonometric polynomial,

$$f_x(t) = 2^{-1/2}x_1 + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + \cdots,$$

containing d terms, and this polynomial is then plotted for $t \in C = [0, 2\pi)$. The mapping is an isometry of R^d onto a subspace of $L^2[0, 2\pi)$ spanned by $\{t \to 2^{-1/2}, t \to \sin t,\ t \to \cos t, \dots\}$. For a fixed t, Andrews notes that the collection of values $\{f_x(t)\}$ as x runs through the data is a projection of the original data onto the unit vector $\gamma(t) = w_d(t)/|w_d(t)|$, where

$$(5.1)\quad w_d(t) = (2^{-1/2}, \sin t, \cos t, \dots, \sin kt) \quad \text{if } d = 2k \text{ is even}$$

$$(5.2)\quad \qquad = (2^{-1/2}, \sin t, \cos t, \dots, \sin kt, \cos kt) \quad \text{if } d = 2k + 1 \text{ is odd.}$$

The Andrews plot can be seen therefore as a projection pursuit method [see, e.g., Friedman and Tukey (1974) and Huber (1985)] as it is used to look for multivariate structure by searching among a (subset) of one-dimensional projections.

More recently, strategies for interactively "touring" through a curve of projections have been described by Asimov and Buja, [see Asimov and Buja (1983), Asimov (1985) and Buja and Asimov (1986)]. For one-dimensional projections, the curve lies in a sphere, for two dimensional projections (ultimately of greater practical interest), the curve lies in a Grassmannian manifold. For simplicity, we confine the analysis to Andrews plots here, though the methods extend, at least in principle, to the newer cases. Asimov (1985) gives bounds and simulations of similar flavor for a number of touring methods, though the tubes approach is not used.

We study the amount of information captured (or missed) by the Andrews plot by asking what fraction of all possible one-dimensional projections are represented. This admittedly crude approach ignores information obtained by going beyond the individual one-dimensional projections to look at the entire two-dimensional plot. However, such "ensemble information" will be harder to extract when the individual projections are presented sequentially and then removed, as in the "grand tour" methods.

Projections of data onto unit vectors u and v that are close will produce similar results—thus it is not necessary (or possible) to look at all projections. Suppose that we deem it unnecessary to use projection directions v which

TABLE 2
Percentage of projections "seen" using an Andrews plot for the curves $\gamma(t)$ in S^{2k} for squint angles 5°(5)20°.

	k					
$\mathbf{2P(W_0 \geq \cos\phi)}$	**1**	**2**	**3**	**4**	**5**	**6**
5°	0.142	0.0019			~0	
10°	0.284	0.0148	0.0006			
15°	0.425	0.0490	0.0046	0.0004		
20°	0.559	0.1132	0.0187	0.0028	0.0004	0.0001

make an angle less than θ with a chosen direction u [Huber (1985) terms this the "squint angle"]. If we employ an Andrews plot, the percentage of possible projections that we see for a squint angle θ is just the ratio of the volume of the tubes of radius $\sqrt{2(1 - \cos\theta)}$ about $\pm\gamma(C)$ to the volume of S^{d-1}. (Both $+u$ and $-u$ are counted since a projection in direction $-u$ is just a reflection of that in direction u). Thus, Hotelling's formula immediately applies.

Consider the case of odd dimensional data, $d = 2k + 1$. The closed curve of projection directions is thus $\gamma(t) = [2/(2k + 1)]^{1/2}$ $(2^{-1/2}, \cos t, \sin t, \dots, \cos kt, \sin kt)$ in S^{2k} for $t \in [0, 2\pi]$. Calculation shows that $|\dot{\gamma}_t|^2 = (k^2 + k)/3$: in particular, γ has constant speed. Thus the length $|\gamma| = 2\pi[(k^2 + k)/3]^{1/2}$. Ignoring overlap (which we may, for the angles listed, as shown below), the percentage of projections seen for a given squint angle θ is given, from (2.3) without caps, by

$$2P(W \geq \cos\theta) = 2\left[(k^2 + k)/3\right]^{1/2} \sin^{2k-1}\theta. \tag{5.3}$$

Squint angles greater than 10° may be unrealistic [indeed, as discussed in Huber (1985), page, 437ff, it is easy to miss the planes produced by the notorious RANDU random number generator]. Thus, Table 2 shows that for even low dimensions ($k = 2, 3$), Andrews plots give only a sparse sampling of the possible projections and suggests that more extensive touring of projections may be needed.

Angles of first overlap. To ensure that the values of Table 2 are exact, we need to know the critical radii θ_c and θ_b at which the tube γ^θ first intersects itself and $-\gamma^\theta$, respectively. A convenient feature of the constant speed curve γ is that $\gamma_s'\gamma_t = [1 + 2\Sigma_1^k \cos j(t - s)]/(2k + 1) = g(t - s)$ is shift invariant.

Thus, θ_c and θ_b can be found by univariate maximization of (4.5) and its two-sided analog. Using a grid on $[0, 2\pi)$ with $30k$ points, the values of $\cot^2\theta_c$ and hence w_c and θ_c were obtained numerically and are displayed in Table 3. The values of ϕ_c hover close to 45°: This is not unexpected since $\gamma_s'\gamma_t$ is close to zero [actually $1/(2k + 1)$] for most values of $s - t$. Hence, the half-angle between γ_s and γ_t is about 45°. From the values of θ_c, one can calculate the volume of the largest tube for which Hotelling's formula (2.3) is exact: This is listed as $P(W \geq w_c)$.

TABLE 3

For curve γ in S^{2k}, θ_ℓ = angle (°) of first local overlap, $w_c = \cos\theta_c$ defines the angle of first global overlap. $P(W \geq w_c)$ = percentage of S^{2k} in largest non-self-overlapping tube, θ_b = angle of first intersection of tubes about γ and $-\gamma$, $2P(W \geq w_b)$ = percentage of S^{2k} in largest nonintersecting tubes.

k	θ_ℓ	w_c	ϕ_c	$P(W \geq w_c)$	θ_b	$2P(W \geq w_b)$
1	54.7	0.578	54.7	0.666	35.3	0.943
2	50.1	0.776	39.2	0.358	37.8	0.653
3	49.1	0.761	40.4	0.230	38.3	0.367
4	48.7	0.757	40.8	0.132	38.5	0.186
5	48.5	0.754	41.0	0.071	38.6	0.091
6	48.4	0.754	41.1	0.037	38.6	0.042
7	48.4	0.753	41.2	0.0189	38.6	0.0189
8	48.3	0.753	41.2	0.0093	38.7	0.0084
9	48.3	0.752	41.2	0.0045	38.7	0.0037
10	48.3	0.752	41.2	0.0022	38.7	0.0016
11	48.3	0.752	41.2	0.0010	38.7	0.0007
12	48.3	0.752	41.3	0.0005	38.7	0.0003
13	48.2	0.751	41.3	0.0002	38.7	0.0001
14	48.2	0.752	41.3	0.0001	38.7	0.0001
15	48.2	0.751	41.3	0.0001	38.7	~0
16	48.2	0.751	41.3	~0	38.7	~0

Using (4.6), we calculate the (constant) squared geodesic curvature of γ as $\{3(3k^2 + 3k - 1)/[5k(k + 1)]\} - 1$, which implies angles θ_ℓ of first local overlap, as shown in Table 3. Aside from the somewhat exceptional case $k = 1$, these are always strictly larger than θ_c, so it follows that nonlocal overlap occurs before (i.e., for smaller tubes than) local overlap.

The values of θ_b for overlap of γ^θ and $-\gamma^\theta$ are similarly obtained and stabilize at 38.7° (as k increases). Since this easily exceeds the tube radii used for Table 2, we conclude that these values are exact.

Quality of approximation when overlap occurs. For tube radii larger than, but close to θ_c, it seems possible that the Hotelling formula (2.4) would continue to provide a useful approximation. We ran a small simulation to investigate this for the curve γ in S^{2k} corresponding to (5.2) For $k = 2(2)12$, we drew 10,000 points U_i from the uniform distribution on S^{2k}. The volume of the tube γ^θ about γ, equivalently $G(\theta) = P(W \geq \cos\theta)$, is estimated by the fraction of $W_i = \sup_t \gamma(t)'U_i$ falling in γ^θ. Representative results for $k = 2, 6, 12$ are given in Table 4 along with values of the Hotelling formula (5.3). Standard errors for $\hat{G}(\theta)$ are estimated by $\{\hat{G}(\theta)(1 - \hat{G}(\theta))/n\}^{1/2}$, so that two standard errors may be conservatively bounded by 0.01.

As seen from Table 3, first overlap occurs in each of the three cases at about 41°. However, Table 4 shows that the Hotelling formula remains accurate for tubes of significantly larger radius as the dimension $d = 2k + 1$ increases. In particular, the formula is good to within 0.01 (also the resolution of the Monte Carlo experiment) for tail probabilities as large as 0.25.

TABLE 4

Columns 2–7: *Simulated* ("sim") *and Hotelling* ["Hot," *cf.* (5.3)] *expressions for probability of tube of geodesic radius* $\theta°$ *about the closed curve* (5.2) *in* S^{2k}. *Horizontal line*: *approximate angle at which first overlap occurs* (cf. *Table* 3.2), *so that* "Hot" *is not exact. Simulations based on* 10,000 *replications, standard error of estimates is at most* 0.005. *Below descending dotted line, Hotelling formula overestimates by at least* 0.01 (*even allowing for simulation uncertainty*). *Columns* 8–11: *extra contribution from the* "*caps*" *term* (2.3).

Angle	$k = 2\ (d = 5)$		$k = 6\ (d = 13)$		$k = 12\ (d = 25)$		caps		
(degrees)	sim	Hot	sim	Hot	sim	Hot	$k = 2$	$k = 6$	$k = 12$
4	0.0	0.001	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.004	0.004	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.013	0.013	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	0.030	0.030	0.0	0.0	0.0	0.0	0.001	0.0	0.0
20	0.056	0.057	0.0	0.0	0.0	0.0	0.003	0.0	0.0
24	0.093	0.095	0.0	0.0	0.0	0.0	0.005	0.0	0.0
28	0.145	0.146	0.001	0.001	0.0	0.0	0.010	0.0	0.0
32	0.211	0.210	0.003	0.004	0.0	0.0	0.016	0.0	0.0
36	0.287	0.287	0.011	0.011	0.0	0.0	0.026	0.0	0.0
40	0.372	0.376	0.031	0.029	0.0	0.0	0.038	0.001	0.0
44	0.459	0.474	0.071	0.068	0.002	0.002	0.054	0.002	0.0
48	0.543	0.580	0.144	0.143	0.008	0.008	0.073	0.004	0.0
52	0.623	0.692	0.266	0.272	0.030	0.030	0.097	0.010	0.0
56	0.691	0.806	0.436	0.476	0.094	0.097	0.124	0.019	0.001
60	0.751	0.919	0.639	0.769	0.253	0.264	0.156	0.034	0.005
64	0.802	1.027	0.805	1.157	0.527	0.620	0.192	0.058	0.013

For this numerical example we conclude, as did Knowles and Siegmund (1989) in a different example, that the Hotelling formula is useful for the range of significance levels encountered in practice, even in those cases where overlap occurs. [Incidentally, the curve γ would arise from a testing problem of the form (2.5) with $\alpha = 0$, $\lambda_i(t) = \cos it$ for $i = 1, \dots, n$, $\lambda_i(t) = \sin(i - n)t$ for $i = n + 1, \dots, 2n$ and $\lambda_0(t) = 2^{-1/2}$.]

Although γ is closed for this example, to obtain a valid upper bound for $P(W \geq \cos\theta)$ for $\theta > \theta_c$, we need to add the caps term in Naiman's bound. The magnitude of this term is shown in the three rightmost columns of Table 4. The term is quite significant for $k = 2$, marginally important for $k = 6$, and negligible for $k = 12$. This is due both to the increased length of γ as k increases and to the smaller relative volume of a cap of fixed angle as k increases.

Most distant projection. Which projection of the data lies furthest from the Andrews plot, and how closely does the plot approach it? Since the projections on γ_t and $-\gamma_t$ are equivalent, this amounts to calculating

$$M = \min_{u \in S^{d-1}} \max_t |u'\gamma_t|,$$

and, if possible, the minimax value of u.

A simple answer exists when $d = 2k + 1$ is odd. From Parseval's identity, for $c = (a_0, a_1, b_1, \dots, a_k, b_k)$,

$$\max_t (c'\gamma_t)^2 \geq \frac{1}{2\pi}\int_0^{2\pi} (c'\gamma_t)^2\, dt = \frac{|c|^2}{2k+1}.$$

Hence, $M \geq 1/\sqrt{2k+1}$, and the bound is attained for $u = (1, 0, \dots, 0)$. For this value of u, the closest projection direction makes an angle $\theta = \cos^{-1}(1/\sqrt{2k+1}) = \pi/2 - (1/\sqrt{2k+1}) + O(k^{-3/2})$. When d is even, the situation is less clear-cut—further details are given in Johansen and Johnstone (1985).

REMARK. The corresponding analysis of $\gamma(t) = w_d(t)/|w_d(t)|$ when d is even is complicated by the fact that $|w_d(t)|^2$ is no longer constant, so that γ has nonconstant speed and $\gamma_t'\gamma_s$ is not shift invariant. This can still be analyzed (for example, using the bivariate methods of Section 4), but a simpler analysis occurs if $w_d(t)$ is modified to $\tilde{w}_d(t) = (\cos t, \sin t, \dots, \cos kt, \sin kt)$, which yields a $\gamma_t = \tilde{w}_d/|\tilde{w}_d|$ of constant speed, etc. The results are qualitatively similar to those described above and are given in detail in Johansen and Johnstone (1985).

6. An application to projection pursuit. Projection pursuit methods form a class of computationally intensive exploratory data analytic procedures aimed at uncovering "nonlinear" structure in multivariate data [see, e.g., Friedman and Tukey (1974), Huber (1985) and Friedman (1987)]. The nature of projection pursuit, whether applied to viewing data, regression, density estimation or classification, is to search over a large number of low dimensional projections in order to optimize numerically a projection index that is sensitive to the particular kind of structure sought.

Such extensive data dredging clearly raises the possibility that spurious structure will be "discovered" [see e.g., Day (1969), Miller (1985) and Friedman (1987)]. It is therefore important to assess the amount of structure that will be found by projection pursuit in white noise. Such significance tests can be based on Monte Carlo simulation, as in Friedman (1987), although this can be enormously expensive computationally.

The purpose of this section is to show that Hotelling's approach (and Weyl's extension) offers, at least in principle, a way to derive approximate tests of significance analytically. We do this by example, working with a slightly idealized model of projection pursuit regression (PPR) in which it is assumed that the independent variables follow a Gaussian distribution. The benefits are that first a simple approximate P-value exists for even a computationally complex procedure such as PPR and, second, a fairly direct application of the Hotelling–Weyl approach can be presented. The price is that further work in less idealized settings is necessary before the approach can be recommended for practical use. Fortunately, Sun (1989) has produced approximate P-values for the practical algorithm of Friedman (1987) (and was actually the first to carry out the program sketched in an earlier version of this paper).

Projection pursuit regression [Friedman and Stuetzle (1981)] fits a regression function of the form $y = \Sigma_k g_k(\alpha_k' x)$ to data $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$, $i = 1, \dots, n$ in an effort to estimate the regression function $f(x)$, the conditional expectation of y given x. If there are several independent variables, the idea is that a small number of *univariate* functions g_k and well-chosen projection directions $\alpha_k \in S^{p-1}$ may lead to parsimonious representation of $f(x)$. [Donoho and Johnstone (1989) have begun some theory to support this heuristic.]

The PPR fit is usually constructed using a least squares criterion: Choose $\{g_k\}$ and $\{\alpha_k\}$ to minimize $\text{ave}\{Y - \Sigma_k g_k(\alpha_k' X)\}^2$, where ave denotes sample average over data points (X_i, Y_i) indexed by $i = 1, \dots, n$. Various algorithms are described in Friedman and Stuetzle (1981) and Friedman (1984). A common algorithm is forward stepwise: First minimize $\text{ave}\{Y - g_1(\alpha_1' X)\}^2$ obtaining a fitted model $Y = \hat{g}_1(\hat{\alpha}_1' X) + e_1$. Repeat the minimization on the residual e_1 obtaining the fit $Y = \hat{g}_1(\hat{\alpha}_1' X) + \hat{g}_2(\hat{\alpha}_2' X) + e_2$, perhaps after a "backfitting" adjustment to $(\hat{g}_1, \hat{\alpha}_1)$ to allow for $(\hat{g}_2, \hat{\alpha}_2)$. The iteration continues until no substantial improvement in fit occurs, as measured by the relative size of $\text{ave}\{\hat{g}_{k+1}^2(\hat{\alpha}_{k+1}' X)\}$ and $\text{ave}\{e_k^2\}$.

A natural approach to significance testing in the stepwise setting is therefore to assess the size of the various terms $\hat{g}(\hat{\alpha}' X)$. We shall consider in detail only the first step, namely, assessment of $\hat{g}_1(\hat{\alpha}_1' x)$. This step is perhaps most important from the viewpoint of significance testing, for if no structure is in fact present, there should be no reason to progress beyond this stage. As is usual with stepwise methods, the results may be formally carried over to later stages, though the distributional assumptions made at the first stage may no longer exactly apply.

We first describe the algorithm by which $\hat{g}_1(\hat{\alpha}_1' x)$ is constructed. In spirit this follows Friedman and Stuetzle (1981) and Huber (1985), but differs in details: Especially, the use of orthogonal polynomials to fit univariate functions is a concession to mathematical convenience. Then we discuss the model assumptions which underlie our theoretical analysis.

Let $e_m(t) = H_m(t)/\sqrt{(m!)}$ denote a normalized Hermite polynomial [e.g., Magnus, Oberhettinger and Soni, 1966] $H_m(t) = e^{t^2/2}(-d/dt)^m e^{-t^2/2}$. For a given direction α, let $g_\alpha(\alpha' x) = \Sigma_{r=1}^m c_r(\alpha) e_r(\alpha' x)$. Let $\bar{Y}$ and s denote the sample mean and standard deviation of $(Y_i)_{i=1}^n$ and standardize the data via $\tilde{Y}_i = s^{-1}(Y_i - \bar{Y})$. The least squares fit is obtained by minimizing the residual sum of squares over $\{c_r(\alpha), r = 1, \dots, m\}$ and α,

$$RSS = \min_{\alpha} \min_{c_r(\alpha)} \sum_{i=1}^n \left\{ \tilde{Y}_i - \sum_{r=1}^m c_r(\alpha) e_r(\alpha' X_i) \right\}^2 .$$

Denote the least squares estimates obtained for fixed α by $\{\hat{c}_r(\alpha)\}$. The statistic

$$T_n^2 = \sum_{i=1}^n \tilde{Y}_i^2 - RSS = \sup_{\alpha} \sum_{i=1}^n \left\{ \sum_{r=1}^m \hat{c}_r(\alpha) e_r(\alpha' X_i) \right\}^2$$

yields a significance test of the null hypothesis of absence of regression, $f \equiv$ constant: reject for large values of T_n^2.

Our idealized model postulates $Y_i = f(X_i) + \varepsilon_i$, with (X_i, Y_i) $i = 1, \dots, n$ i.i.d. and with errors ε_i having mean zero and variance σ_ε^2 and being independent of $X_i \sim N_p(0, I)$. We shall illustrate the Hotelling–Weyl approach in deriving the following.

THEOREM 6.1. *Let* $b_{p,m} = E[\Sigma_1^m rV_r^2]^{(p-1)/2}$ *for* $(V_1, \dots, V_m)$ *a uniformly distributed random vector on* S^{m-1}. *Then, under the null hypothesis that* $f \equiv$ *constant, as* $x \to \infty$,

$$\lim_{n\to\infty} P\left(T_n^2 \ge x^2\right) \sim \frac{\omega_{p-1}\omega_{m-1}}{\omega_{p+m-2}} b_{p,m} P\left(\chi_{p+m-1}^2 \ge x^2\right) \tag{6.1}$$

$[\omega_{d-1} = 2\pi^{d/2}/\Gamma(d/2)]$.

The constant $b_{p,m}$ can, in principle, be evaluated exactly when p is odd [e.g., if $p = 3$, $b_{p,m} = (m + 1)/2$], and is in any case straightforward to estimate by simulation.

The demonstration breaks into three steps:

1°. The random variable T_n converges in distribution to a random variable T with representation

$$T = \max_{\alpha \in S^{p-1}} \max_{\beta \in S^{m-1}} \gamma(\alpha, \beta)'Z, \tag{6.2}$$

where $Z \sim N_d(0, I)$ and γ: $S^{p-1} \times S^{m-1} \to S^{d-1}$ is a smooth mapping to be described below. This representation depends critically on an idea suggested by David Siegmund.

2°. Decomposing $Z = RU$, where $R \sim \chi_{(d)}$ and $U \sim \text{Uniform}(S^{d-1})$ are independent, we write $T = RW$, where in analogy with (2.2), $W = \sup_{\gamma \in C} \gamma' U$, where $C = \gamma(S^{p-1} \times S^{m-1})$ is now a *surface* in S^{d-1} of dimension $p + m - 2$ rather than a curve. From the decomposition

$$P(T \ge x) = \int_x^\infty P(W \ge xr^{-1}) P(R \in dr),$$

we are led, exactly as in the case of curves, to consider the volume (in S^{d-1}) of a tube about the surface C. Weyl's theorem, specialized as in Corollary 2.5 yields

$$\begin{aligned} P(T \ge x) &\sim \frac{k_0}{\omega_\kappa} \int_x^\infty P\left[\left(U_1^2 + \cdots + U_{\kappa+1}^2\right)^{1/2} \ge \frac{x}{R} \,\middle|\, R = r\right] P(R \in dr) \qquad (x \to \infty) \\ &= \frac{k_0}{\omega_\kappa} P\left(\chi_{(\kappa+1)}^2 \ge x^2\right), \end{aligned} \tag{6.3}$$

where $\kappa = p + m - 2$ is the dimension and k_0 is the volume of the manifold C. [Inasmuch as the manifold C has no boundary, this example is simpler than the statistical applications of Weyl's theorem given by Naiman (1990) and Knowles and Siegmund (1989).]

3°. A calculation shows that the volume $k_0 = \omega_{p-1}\omega_{m-1}E(\sum_1^m rV_r^2)^{(p-1)/2}$, where $(V_1, \dots, V_m)$ is a uniform random vector on S^{m-1}. Combining this with (6.3) yields the formula (6.1).

We proceed to the details of steps 1° and 3°.

1°. To approximate the null distribution of T_n, we separate α from the data X_i by using the Hermite polynomial addition theorem [e.g., Magnus, Oberhettinger and Soni (1966), page 254],

$$e_r(\alpha' x) = \sum_{|k|=r} \sqrt{\binom{r}{k}} \alpha^k e_k(x). \tag{6.4}$$

Here we use multiindex notation: $k = (k_1, \dots, k_p)$, $\alpha^k = \alpha_1^{k_1} \cdots \alpha_p^{k_p}$, $|k| = k_1 + \cdots + k_p$ and $e_k(x) = e_{k_1}(x_1) \cdots e_{k_p}(x_p)$. The generalized binomial coefficient $\binom{r}{k} = r!/\prod k_i!$. The summation is taken over the $d_r = d(r, p) = \binom{r+p-1}{p-1}$ vectors $(k_1, \dots, k_p)$ for which $k_1 + \cdots + k_p = r$.

Define the $n \times d$ matrix X by $X_{ik} = e_k(X_i)$ and the $d \times m$ matrix $\Gamma(\alpha)$ by

$$\Gamma_{kr}(\alpha) = \begin{cases} \sqrt{\binom{r}{k}} \alpha^k & \text{if } |k| = r, \\ 0 & \text{otherwise.} \end{cases}$$

Let $c(\alpha) = (c_r(\alpha))_{r=1}^m$. The projection pursuit least squares minimization problem becomes, in matrix notation, to minimize $|\tilde{Y} - X\Gamma(\alpha)c(\alpha)|^2$ over α and $c(\alpha)$. Hence, $\hat{c}(\alpha) = (\Gamma(\alpha)'X'X\Gamma(\alpha))^{-1}\Gamma(\alpha)'X'\tilde{Y}$ and

$$T_n^2 = \max_\alpha \tilde{Y}'X\Gamma(\alpha)(\Gamma(\alpha)'X'X\Gamma(\alpha))^{-1}\Gamma(\alpha)'X'\tilde{Y}.$$

Let $Z_n = n^{-1/2}X'\tilde{Y}$ and $E_n = n^{-1}X'X$. It follows that

$$T_n^2 = \max_\alpha Z_n'\Gamma(\alpha)(\Gamma(\alpha)'E_n\Gamma(\alpha))^{-1}\Gamma(\alpha)'Z_n.$$

Under H_0, Z_n has mean 0 and (from the orthogonality of the Hermite polynomials) asymptotic covariance matrix I_d, and so the central limit theorem guarantees that Z_n converges in distribution to a standard Gaussian vector Z. Again by the orthogonality of the Hermite polynomials, the law of large numbers guarantees convergence of E_n in probability under H_0 to the identity matrix I_d. Hence, (Z_n, E_n) converges jointly in distribution to (Z, I). The function $(z, e) \to \max\{z'\Gamma(\alpha)(\Gamma(\alpha)'e\Gamma(\alpha))^{-1}\Gamma(\alpha)'z,\ \alpha \in S^{p-1}\}$ is continuous for $z \in R^n$ and e in a neighborhood of I_m, and so the continuous mapping theorem ensures that T_n^2 converges in distribution to the maximum of a chi-square process,

$$T^2 = \max_\alpha Z'\Gamma(\alpha)\Gamma(\alpha)'Z,$$

where we have used the easily checked fact that $\Gamma(\alpha)'\Gamma(\alpha) = I$.

Now introduce $d_r \times 1$ vectors $\gamma^r(\alpha)$ defined by

$$\gamma_k^r(\alpha) = \sqrt{\binom{r}{k}}\alpha^k \quad \text{for } |k| = r.$$

Write Z^r for the $d_r \times 1$ component of Z corresponding to subscripts k with $|k| = r$. From the definition of $\Gamma(\alpha)$,

$$T^2 = \max_{\alpha} \sum_{r=1}^{m} (\gamma^r(\alpha)'Z^r)^2.$$

Siegmund's idea is to represent the maximum of the chi-square process in terms of the maximum of a Gaussian process using the device

$$\left[\sum_{r=1}^{m} (\gamma^{r\prime}z^r)^2\right]^{1/2} = \sup_{\beta \in S^{m-1}} \sum_{r=1}^{m} \beta_r(\gamma^{r\prime}z^r), \qquad z^r \in \mathbb{R}^{d_r}.$$

Thus,

$$T = \sup_{\alpha \in S^{p-1}} \sup_{\beta \in S^{m-1}} \sum_{r=1}^{m} \beta_r(\gamma^r(\alpha)'Z^r) = \sup_{\alpha,\beta} \gamma(\alpha,\beta)'Z,$$

which is a representation of the form (6.2) with

$$(6.5) \qquad \gamma(\alpha,\beta) = \begin{pmatrix} \beta_1\gamma^1(\alpha) \\ \vdots \\ \beta_m\gamma^m(\alpha) \end{pmatrix} \in S^{d-1}.$$

For future reference, we record some properties of the mapping γ^r: $S^{p-1} \to S^{d_r-1}$. The identity $|\gamma^r|^2 = 1$ reflects the fact that the total mass of the multinomial distribution of the vector $K = (K_1, \dots, K_p)$ counting the distribution of r balls into p cells with probabilities $(\alpha_1^2, \dots, \alpha_p^2)$ equals 1. We compute partial derivatives

$$D_{\alpha_j}\gamma_k^r = \sqrt{\binom{r}{k}} k_j \alpha^{k-e_j},$$

and hence from the properties of multinomial distribution

$$(6.6) \qquad (D_{\alpha_j}\gamma^r)'(D_{\alpha_{j'}}\gamma^r) = \frac{1}{\alpha_j\alpha_{j'}} EK_jK_{j'} = r\delta_{jj'} + r(r-1)\alpha_j\alpha_{j'}.$$

3°. Finally, we compute the volume of $C = \gamma(S^{p-1} \times S^{m-1})$ by expressing the volume element on C in terms of the volume element $d\sigma_{p-1}(\alpha)\, d\sigma_{m-1}(\beta)$ on $S^{p-1} \times S^{m-1}$. Let ψ: $U \subset \mathbb{R}^{p+m-2} \to C$, ϕ: $U_1 \times U_2 \subset \mathbb{R}^{p-1} \times \mathbb{R}^{m-1} \to S^{p-1} \times S^{m-1}$ be local coordinates ("charts") on C and $S^{p-1} \times S^{m-1}$, respectively (U, U_1, U_2 are open sets), that are connected by γ: $\psi = \gamma \circ \phi$. Using D to denote differential (Jacobian matrix), the volume element on C in local coordinates becomes

$$|D\psi'\, D\psi|^{1/2} = |D\phi'\, D\gamma'\, D\gamma\, D\phi|^{1/2}.$$

Writing $\gamma_j^r = D_{\alpha_j}\gamma^r(\alpha)$ and partitioning according to $\alpha \in S^{p-1}$, $\beta \in S^{m-1}$, we find from (6.5) that

$$D\gamma = \left[D_\alpha \gamma | D_\beta \gamma\right] = \left[\begin{array}{ccc|ccc} \beta_1 \gamma_1^1 & \cdots & \beta_1 \gamma_p^1 & \gamma^1 & & 0 \\ \vdots & & \vdots & & \ddots & \\ \beta_m \gamma_1^m & \cdots & \beta_m \gamma_p^m & 0 & & \gamma^m \end{array}\right].$$

An appeal to (6.6) yields

$$D\gamma' D\gamma = \begin{bmatrix} a(\beta) I_p + b(\beta)\alpha\alpha' & \alpha\tilde{\beta}' \\ \tilde{\beta}\alpha' & I_m \end{bmatrix}$$

where $\tilde{\beta}_r = r\beta_r$, $a(\beta) = \sum_1^m r\beta_r^2$ and $b(\beta) = \sum_1^m r(r-1)\beta_r^2$. Writing $\phi(s,t) = (\alpha(s), \beta(t))$, we get

$$D\phi = \begin{bmatrix} D\alpha & 0 \\ 0 & D\beta \end{bmatrix}$$

and (since $|\alpha|^2 = 1 \Rightarrow \alpha' D\alpha = 0$)

$$|D\psi' D\psi|^{1/2} = a^{(p-1)/2}(\beta) |D\alpha' D\alpha|^{1/2} |D\beta' D\beta|^{1/2}.$$

Hence,

$$\begin{aligned} \operatorname{vol}(C) &= \int_{S^{p-1} \times S^{m-1}} a^{(p-1)/2}(\beta)\, d\sigma_{p-1}(\alpha)\, d\sigma_{m-1}(\beta) \\ &= \omega_{p-1}\omega_{m-1} E\left(\sum_1^m r W_r^2\right)^{(p-1)/2}, \end{aligned}$$

where $W_1, \dots, W_m$ are uniformly distributed on S^{m-1}.

Acknowledgments. This paper, first drafted in April 1985, was stimulated by a 1983 seminar at Stanford on Hotelling's work and a summer 1984 IMS meeting talk by R. A. Olshen on his work in gait analysis. The authors thank Richard Olshen for stimulating discussions and hospitality, and Persi Diaconis, Jerry Friedman, Daniel Naiman, Mehrdad Shahshahani and David Siegmund for many helpful conversations. Thanks go especially to Ed Biden for providing the gait data used in Section 3. We are also grateful to the referees and Associate Editor for their extensive and stimulating comments throughout, and regarding Section 6 in particular.

The first author's work was supported in part by NIH grant PMS CA 26666 and the second author's in part by NSF grants MCS 80-24649, 84-51750, 86-00235, 87-14169, NIH grant PHS GM 21215-12 and the Sloan Foundation.

REFERENCES

ANDREWS, D. F. (1972). Plots of high-dimensional data. *Biometrics* **28** 137–156.

ASIMOV, D. (1985). The Grand Tour: A tool for viewing multidimensional data. *SIAM J Sci. Statist. Comput.* **85** 128–143

ASIMOV, D. and BUJA, A. (1983). Finding structure in unstructured data (a short film). Computation Research Group, Stanford Linear Accelerator Center.

Bradley, R. A. (1952). Corrections for non-normality in the use of two-sample t and F tests at high significance levels. *Ann. Math. Statist.* **23** 103.

Buja, A. and Asimov, D. (1986). Grand Tour methods: An outline. In *Computer Science and Statistics: Proceedings of the 17th Symposium on the Interface* (D. M. Allen, ed.) 63–67, North-Holland, Amsterdam.

Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika* **56** 463–474.

Diaconis, P. and Efron, B. (1985). Testing for independence in a two-way table: New interpretations of the chi-square statistic (with discussion). *Ann. Statist.* **13** 845–913.

Donoho, D. L. and Johnstone, I. M. (1989). Projection based smoothing, and a duality with kernel methods. *Ann. Statist.* **17** 58–106.

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

Friedman, J. H. (1984). SMART (smooth multiple additive regression technique): User's guide. Technical Report, Dept. Statist., Stanford Univ.

Friedman, J. H. (1987). Exploratory projection pursuit. *J. Amer. Statist. Assoc.* **82** 249–266.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

Friedman, J. H. and Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.* **C-23** 881–889.

Hastie, T. and Stuetzle, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84** 502–516.

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

Huber, P. J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13** 435–525.

Johansen, S. (1984). *Functional Equations, Random Coefficients and Nonlinear Regression with Application to Kinetic Data.* Springer, Berlin.

Johansen, S. and Johnstone, I. M. (1985). Some uses of spherical geometry in simultaneous inference and data analysis. Technical Report No. 237, Dept. Statist., Stanford Univ.

Johnson, L. W. (1977). Stochastic parameter regression: An annotated bibliography. *Internat. Statist. Rev.* **45** 257–272.

Johnson, L. W. (1980). Stochastic parameter regression: An additional annotated bibliography. *Internat. Statist. Rev.* **48** 95–102.

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

Keeping, E. S. (1951). A significance test for exponential regression. *Ann. Math. Statist.* **22** 180–198.

Knafl, G., Sacks, J. and Ylvisaker, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80** 683–691.

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.

Magnus, W., Oberhettinger, F. and Soni, R. P. (1966). *Formulas and Theorems for the Special Functions of Mathematical Physics.* Springer, New York.

Miller, R. G., Jr. (1981). *Simultaneous Statistical Inference.* Springer, New York.

Miller, R. G., Jr. (1985). Discussion of "Projection Pursuit" by Peter J. Huber, *Ann. Statist.* **13** 510–513.

Mulholland, H. P. (1965). On the degree of smoothness and on singularities in distributions of statistical functions. *Proc. Cambridge Philos. Soc.* **61** 721–739.

Mulholland, H. P. (1970). On singularities of sampling distributions, in particular for ratios of quadratic forms. *Biometrika* **57** 155–174.

Naiman, D. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.* **14** 896–906.

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

OLSHEN, R. A., BIDEN, E. N., WYATT, M. P. and SUTHERLAND, D. H. (1989). Gait analysis and the bootstrap. *Ann. Statist.* **17** 1419–1440.

RAO, C. R. (1965). The theory of least squares when the parameters are stochastic and its applications to the analysis of growth curves. *Biometrika* **52** 447–458.

SANTALÓ, L. A. (1976). *Integral Geometry and Geometric Probability.* Addison-Wesley, Reading, Mass.

SIDDIQUI, M. M. (1958). Distribution of a serial correlation coefficient near the ends of the range. *Ann. Math. Statist.* **29** 852–861.

SPJØTVOLL, E. (1977). Random coefficients regression models. A review. *Math. Operationsforsch. Statist. Ser. Statist.* **8** 69–93.

SUN, J. (1989). P-values in projection pursuit. Technical Report No. 104, Dept. Statist., Stanford Univ.

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

WYNN, H. P. and BLOOMFIELD, P. (1971). Simultaneous confidence bands in regression analysis. *J. Roy. Statist. Soc. Ser. B* **33** 202–217.

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN
5 UNIVERSITETSPARKEN
DK-2100 COPENHAGEN Ø
DENMARK

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305