SIAM J. SCI. STAT. COMPUT.
Vol. 6, No. 1, January 1985

011

THE GRAND TOUR: A TOOL FOR VIEWING MULTIDIMENSIONAL DATA*

DANIEL ASIMOV†

Abstract. The grand tour is a method for viewing multivariate statistical data via orthogonal projections onto a sequence of two-dimensional subspaces. The sequence of subspaces is chosen so that it is dense in the set of all two-dimensional subspaces. Desirable properties of such sequences of subspaces are considered, and several specific types of sequences are tested for rapidity of becoming dense. Tabulations are provided of the minimum length of a grand tour sequence necessary to achieve various degrees of denseness in dimensions up to 20.

Key words. multivariate data, multidimensional data, exploratory data analysis, computer graphics, scatterplots, Andrews plot, grand tour

1. Introduction. The familiar "scatterplot" (of a finite sample of ordered pairs of variables) can be extraordinarily informative. Thus, it is very tempting to consider the p-dimensional scatterplot—a finite sample of ordered p-tuples of variables—and to devise ways to view it.

Even for $p=3$, we have no magic pen that draws points in mid-air. Resorting to computer graphics [FFT], however, will permit us to see the three-dimensional scatterplot on a display screen just as if the points were drawn in mid-air. With the aid of a graphical input device like a "trackball," we may even rotate the scatterplot in real time.

For p greater than 3, we are faced with serious problems. How can computer graphics technology be used, in conjunction with our visual abilities, to better grasp the structure of the p-dimensional data?

A simple answer to this question is to project the data orthogonally onto some two-dimensional subspace of p-dimensional Euclidean space, and then to view the resulting projected image.

A problem immediately arises: Which of the infinitely many two-dimensional subspaces shall we choose for viewing? The idea of the grand tour is to move through a sequence of projections, chosen to be dense in the set of all projections. As a result, we can view (or else have the computer apply some analysis or measurement to) a sequence of two-dimensional scatterplots which, asymptotically, come arbitrarily close to *all* 2-dimensional scatterplots projectable from the given data.

Historically, the grand tour is a descendant of the Andrews plot [Andr] which dates to 1972. This plot is often realized as a stationary set of function graphs $y=f_i(t)$ where $f_i(t)=x_1/\sqrt{2}+x_2\sin t+x_3\cos t+x_4\sin 2t+x_5\cos 2t+\cdots$ for the ith data point $(x_1, x_2, x_3, \cdots, x_p)$. This can, however, be interpreted also as a time sequence $\{f_1(t), \cdots, f_N(t)\}$ of points in R, where at time t_0 we are viewing the dot-products of all the data points with the vector given by $(1/\sqrt{2}, \sin t_0, \cos t_0, \sin 2t_0, \cos 2t_0, \cdots)$.

Then, in 1977, Paul and John Tukey [TT77] presented some further thoughts on Andrews plots, including an example of a dense curve of directions in R^4. They also considered briefly a two-dimensional (not necessarily dense) version of Andrews plots which they called "ouija" plots.

* Received by the editors November 8, 1983. This work was supported by the U.S. Department of Energy under contracts DE-AC03-76SF00515 and DE-AT03-81-ER10843.

† Department of Computer Science, University of California, Berkeley, California 94720.

Meanwhile, real-time computer graphical visualization of three-dimensional (or higher) rotation had been achieved when the PRIM-9 system was implemented at the Stanford Linear Accelerator Center (SLAC) in the early 1970's [FFT].

Much of the work described herein was performed at Harvard University in 1980–81 and at SLAC in 1981–83.

2. Overview. In order to implement a grand tour on a computer graphics system, it is necessary to have an explicitly computable sequence of orthonormal 2-frames (a 2-frame is an orthonormal pair of vectors) in p-dimensional Euclidean space. The p-dimensional data is then projected, in turn, onto the 2-plane[1] spanned by each 2-frame. If desired, each projected image may be displayed on the screen, or else processed somehow by the computer (or both). We list below some desiderata for this sequence of 2-frames:

Desiderata. A) The sequence of planes should be *dense* in the space of all planes. Precisely, let $G_{2,,p}$ stand for the space of unoriented 2-planes through the origin in Euclidean p-space (a so-called "Grassmannian manifold"). Let $P_1, P_2, \cdots$ be the infinite sequence of 2-planes (spanned by the infinite sequence of 2-frames generating the grand tour). Then our condition A says that for every 2-plane P and for every $\varepsilon > 0$, there exists an n such that the distance $d(P, P_n)$ from P to P_n is less than ε. (Our definition of the distance function d is in § 4.) Note that this denseness is not just a desideratum, but part of our definition of "grand tour."

B) Our sequence of planes should become dense in $G_{2,p}$ *rapidly.* This means finding an efficient algorithm to compute the sequence of 2-frames and to project the p-dimensional date onto each pair of vectors in turn.

C) It would be useful for the sequence of planes to be uniformly distributed in $G_{2,p}$. That is to say, for each open measurable subset A of $G_{2,p}$, our sequence of planes $P_1, P_2 \cdots$ should pass through A with frequency proportional to the measure of A. We refer here to the invariant measure μ on $G_{2,p}$ (which is uniquely determined up to a positive constant factor). Precisely, we want

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_A(P_i) = \mu(A),$$

where I_A is the characteristic function of the set A.

D) Our sequence of planes should be continuous, in some sense, if its projections are to be apprehended by a human observer. Each plane should be perceptibly close to those planes just before and after it in the sequence. (This condition is of no importance in many applications of the grand tour in which no human observing occurs.)

E) For human observers, our sequence of planes should be as straight as possible. That is, if we think of the planes as being evenly-spaced points on a curve in $G_{2,p}$, then we should be able to choose that curve so that it is almost a geodesic. This is another way of assuring that the sequence of planes is both comprehensible to the observer, and also that it moves rapidly to new views, giving new information about the data being projected.

F) The grand tour ought to have a built-in degree of flexibility about it. This would enable the user to better optimize those qualities (among A) through E, for example) which may be important for the particular purpose he or she has in mind. Flexibility may be obtained by finding a parametric family of sequences of planes. There should then be some clear relationship between the parameter(s) and the desired

[1] *Note.* Unless otherwise specified, all "planes" referred to herein will be planes through the origin.

properties, so that the user can choose the parameter(s) wisely. It should also be possible to interactively change parameters after the grand tour has begun.

G) The sequence of planes should be reconstructible at any later occasion. In practice, this simply means that either the sequence of planes is chosen from a parametric family with parameters known to the user, or else there may be a pseudo-random component whose random number algorithm(s) and seeds(s) are known. It is, of course, desirable that in reconstructing a particular plane of our sequence, the other planes preceding it need not be computed all over again.

Remarks. R1. To require bona fide denseness of the infinite sequence of planes $P_1, P_2, \cdots$ is unnecessary for any real-world implementation of the grand tour. In particular, if we know in advance the number L of planes $P_1, P_2, \cdots, P_L$ we will be using, we can dispense entirely with the idea of an infinite sequence. We may also be able to better optimize the seven properties A) through G) once L is known.

R2. If the method for producing the sequence $P_1, P_2, \cdots$ is based on some random process, we will generally be able to claim properties such as denseness or uniformity as being "almost sure" rather than certain. (But this is almost surely sufficient for our purposes!)

R3. There is evidently a tradeoff between rapidity and continuity. This suggests using a *curve* of points in $G_{2,p}$, and obtaining a sequence $P_1, P_2, \cdots$ by walking along this curve after choosing an appropriate stepsize. The human observer, of course, desires continuity. A machine alone, processing many planes, will rather need rapidity.

R4. To date, the known sequences $P_1, P_2, \cdots$ that are both *uniform* and *rapid* require sequences of pseudo-random numbers to compute them. Thus to achieve reconstructability the algorithm and seed value of the pseudo-random sequence must be retained.

R5. Minor violations of uniformity are acceptable for the human observer. Strict uniformity is needed only when the computer is determining distributional properties of some statistic of two-dimensional scatterplots (see § 5).

R6. In all of the above, we have emphasized the choice of 2-planes $P_1, P_2, \cdots$. In practice, however, when displaying a two-dimensional picture we must also choose its rotational position on the screen. This is accomplished by choosing not just a mere plane P_i but rather a pair of orthonormal vectors $(\mathbf{v}_i, \mathbf{w}_i)$ spanning P_i: these are to be identified with the X and Y directions of the display screen. Even when no display is required, the computer must still hold internally a description of each plane P_i. The "2-frame" $(\mathbf{v}_i, \mathbf{w}_i)$ is a convenient form for this information.

3. Some specific grand tours. In this section, we present three general methods for producing grand tours.

I. Torus method. The N-dimensional torus T^N may be defined as the Cartesian product of N identical unit circles. Equivalently, T^N may be thought of as Euclidean space R^N in which all arithmetic is performed modulo one. Symbolically, $T^N \simeq R^N/(2\pi Z^N)$ where Z^N is the integer lattice in R^N. It is well-known that dense curves may be found on T^N via the following.

PROPOSITION [HWTN]. *Let $\{\lambda_1, \cdots, \lambda_N\}$ be a set of real numbers that are linearly independent over the integers.*[2] *Then the curve $\alpha: R \to T^N$ via $\alpha(t) = (\lambda_1 t, \cdots, \lambda_N t)$ has dense image in T^N. (Note that the coordinates $\lambda_i t$ are interpreted modulo 2π.)*

[2] Real numbers $u_1, \cdots, u_N$ are said to be *linearly independent over the integers* if the only sequence of integers $\{K_1, \cdots, K_N\}$ for which the equation $\sum_{i=1}^N K_i u_i = 0$ holds is with all $K_i = 0$.

The special orthogonal group in dimension p, denoted $SO(p)$, is the set of all orthogonal $p\times p$ matrices having determinant $=+1$. $SO(p)$ has a topology induced from R^{p^2}, the space of all real $p\times p$ matrices, and is in this way a compact manifold of dimension $\frac{1}{2}(p^2-p)$. $SO(p)$ may equivalently be thought of as the space of all rotations of the unit sphere in R^p. (As such, it is a "Lie group" [Chev].)

We let $R_{ij}(\theta)$ denote the element of $SO(p)$ which rotates the standard basis vector $\mathbf{e}_i$ through an angle θ towards the standard basis vector $\mathbf{e}_j$ inside the i, j coordinate 2-plane of R^p, leaving fixed the orthogonal complement of this 2-plane.

We let $G_{2,p}$ denote the space of all 2-planes in R^p. We also let $V_{2,p}$ denote the space of all ordered pairs of orthonormal vectors in R^p. ($V_{2,p}$ is topologized as a subset of $R^p\times R^p$ and is compact.) We have the natural continuous surjections $\pi: SO(p)\to V_{2,p}$ and $\rho: V_{2,p}\to G_{2,p}$ given by $\pi(Q)=(Q\mathbf{e}_1, Q\mathbf{e}_2)$ for any $Q\in SO(p)$, and $\rho(\mathbf{v},\mathbf{w})=$ the 2-plane spanned by $\mathbf{v}$ and $\mathbf{w}$, for any $(\mathbf{v},\mathbf{w})\in V_{2,p}$.

We are now ready to describe explicitly the torus method.

1. Let $N=\frac{1}{2}(p^2-p)$ and think of the coordinates of T^N as being indexed by all pairs i, j with $1\leqq i<j\leqq p$.

2. Define a map $f: T^N\to SO(p)$ via

$$f(x_{1,2},\cdots,x_{p-1,p})=R_{12}(x_{1,2})\circ\cdots\circ R_{p-1,p}(x_{p-1,p}).$$

In words: f is the product of coordinate-plane rotations through angles determined by the toral coordinates. (Note: each x_{ij} is only well defined modulo 2π, but since $R_{ij}(\theta+2\pi)=R_{ij}(\theta)$, f is well defined.)

3. We claim that f is a surjection. This fact was in essence discovered by L. Euler [MMCM]; the angles $\{x_{ij}\}$ are referred to as "Euler angles."

4. Choose real numbers $\lambda_1,\cdots,\lambda_N$ and a stepsize STEP such that the numbers $\{2\pi, \mathrm{STEP}\cdot\lambda_1,\cdots,\mathrm{STEP}\cdot\lambda_N\}$ are linearly independent over the integers. Use $\lambda_1,\cdots,\lambda_N$ to define the curve $\alpha: R\to T^N$ via $\alpha(t)=(\lambda_1 t,\cdots,\lambda_N t)$ as in Proposition 1. Thus, we know that the image $\alpha(R)$ of α is dense in T^N.

5. We conclude, therefore that $f\circ\alpha: R\to SO(p)$ has dense image $f(\alpha(R))$ in $SO(p)$.

6. The discrete sequence $\{f\circ\alpha(K\cdot\mathrm{STEP}), K=1,2,\cdots\}$ must therefore also be dense in $SO(p)$.

7. Finally, we define our sequence of 2-frames $(\mathbf{v}_K,\mathbf{w}_K)$ as $(\mathbf{v}_K,\mathbf{w}_K)=\pi\circ f\circ\alpha(K\cdot\mathrm{STEP}), K=1,2,\cdots$. By 6 above, this sequence must be dense in $V_{2,p}$.

8. We define our sequence of 2-planes $P_1, P_2,\cdots$ as, of course, $P_K=\rho(V_K, W_K)=\rho\circ\pi\circ f\circ\alpha(K\cdot\mathrm{STEP})$.

It follows from 7 above that this sequence is dense in $G_{2,p}$. This concludes our description of how to compute a grand tour by the torus method.

Remarks. R1. The number N, the dimension of the torus used here, can be reduced from $\frac{1}{2}(p^2-p)$ to $2p-3$ (see Appendix). The resulting sequence of orthogonal matrices will no longer be dense in $SO(p)$ but will be dense when pushed via π and ρ into $V_{2,p}$ and $G_{2,p}$. This reduction achieves a considerable savings in computation time.

R2. The sequence given by $\mathbf{z}_K=(K\cdot\mathrm{STEP}\cdot\lambda_1,\cdots,K\cdot\mathrm{STEP}\cdot\lambda_N)\in T^N$ is uniformly distributed on T^N. But the maps f, π, and ρ do not respect volumes. Thus, the sequences of 2-frames $\{(\mathbf{v}_K,\mathbf{w}_K)\}$ and planes $\{P_K\}$ are *not* uniformly distributed. This remark applies equally to the $2p-3$ version in the Appendix.

R3. The parameter STEP may be varied before, or even during, each run of the grand tour. The effect of increasing the size of STEP is to trade continuity for rapidity. More accurately, this is true for some range of values $0<\mathrm{STEP}\leqq M$, after which there is very little noticeable effect of STEP on either continuity (which is totally lost) or rapidity (which is at a maximal level).

R4. Although it is convenient to fix the values of $\lambda_1, \cdots, \lambda_N$ and vary STEP, it is in fact the vector $\mathbf{x} = (\mathrm{STEP} \cdot \lambda_1, \cdots, \mathrm{STEP} \cdot \lambda_N)$ in the torus T^N which determines the characteristics of the grand tour, torus method. If the total number L of planes to be used is known, then Korobov [MCTP] has determined vectors $\mathbf{x}$ which behave optimally vis-a-vis the distribution of the sequence $\mathbf{x}, 2 \cdot \mathbf{x}, 3 \cdot \mathbf{x}, \cdots$ in T^N. It seems likely that Korobov coefficients will give rise to sequences of 2-frames and 2-planes which become dense rapidly, but their use is restricted to occasions when L is known in advance. Alternatively, some easy-to-compute values of $\mathbf{x}$ seem to work very well. For example, two choices are

a) Let $\lambda_K = \sqrt{p_K}$ = the square root of the Kth prime ($p_1 = 2, p_2 = 3, \cdots$). Let STEP = almost any irrational positive real.

b) Let $\lambda_K = e^K \bmod 1$ ($e = 2.71828 \cdots$) and again let STEP = almost any irrational positive real.

II. At-random method. In this method, each 2-frame is chosen independently, from the "uniform" distribution on $V_{2,p}$. This distribution is more accurately termed the "invariant" measure on $V_{2,p}$, because it is characterized up to constant factor by its invariance under the action of $SO(p)$ on $V_{2,p}$. That is, if $Q \in SO(p)$ and $A \subset V_{2,p}$, then we have for the invariant measure m,

$$m(A) = m\{(Q\mathbf{v}, Q\mathbf{w})(\mathbf{v}, \mathbf{w}) \in A\}.$$

To pick the sequence $\{(\mathbf{v}_K, \mathbf{w}_K)\}$ of 2-frames, we use the "rejection" method as follows:

1. Generate a sequence of pseudorandom numbers $x_1, x_2, \cdots$ in the unit interval.

2. Set $y_1 = 2x_1 - 1$, $y_2 = 2x_2 - 1, \cdots$.

3. Assume we have already used the random numbers $y_1, \cdots, y_n$ (at the start $n = 0$). Set $z_i = y_{n+i}$ for $i = 1, \cdots, p$.

4. Test for $0 < z_1^2 + \cdots + z_p^2 \leqq 1.0$. If not, return to Step 3 and try again.

5. Go through Step 3 again until a second set of p numbers are found (call them $u_1, \cdots, u_p$ this time) with $0 < u_1^2 + \cdots + u_p^2 \leqq 1.0$.

6. Letting $\mathbf{z} = (z_1, \cdots, z_p)$ and $\mathbf{u} = (u_1, \cdots, u_p)$, apply the Gram–Schmidt procedure to obtain an orthonormal pair of vectors

$$\mathbf{v}_K = \mathbf{z}/\|\mathbf{z}\| \quad \text{and} \quad \mathbf{w}_K = \frac{\mathbf{u} - (\mathbf{u} \cdot \mathbf{v}_K)\mathbf{v}_K}{\|\mathbf{u} - (\mathbf{u} \cdot \mathbf{v}_K)\mathbf{v}_K\|}.$$

These constitute the next 2-frame $(\mathbf{v}_K, \mathbf{v}_K)$ of our sequence. It is easy to verify that despite the apparent asymmetry in the use of the Gram–Schmidt procedure, $(\mathbf{v}_K, \mathbf{w}_K)$ is in fact selected at random from the invariant distribution.

7. It follows immediately that the corresponding sequence of planes $P_K = \langle \mathbf{v}_K, \mathbf{w}_K \rangle$ may be thought of as being selected from the corresponding invariant distribution on $G_{2,p}$.

Remarks. R1. The at-random method has in its favor the extreme ease of concept and computation. It is too discontinuous (totally) for movie viewing. (This is, of course, no problem if the viewer prefers to see only a sequence of still pictures.)

R2. The at-random method will produce, almost surely, a uniformly distributed sequence.

R3. There is no flexibility in the at-random method.

R4. The at-random method becomes dense about as fast as the torus method with large stepsize.

III. Random-walk method. The random-walk method was devised in an attempt to unite the flexibility of the torus method with the guaranteed uniform distribution of the at-random method. We describe here two methods, the plain random walk and the smoother random walk.

A. *The plain random walk.* Let μ denote a measure on $SO(p)$ satisfying the following condition:

Condition D. The support of μ (i.e., the complement of the union of all open μ-null sets) generates a dense subgroup of $SO(p)$. Then we obtain a sequence of orthogonal matrices $Q_K \in SO(p)$ as follows:

1. Set $Q_0 = I_p$, the identity matrix.
2. For $K = 1, 2, \cdots$ we let $g_1, g_2, \cdots$ be selected i.i.d., according to the law μ, from $SO(p)$.
3. For $K = 1, 2, \cdots$ we set $Q_K = g_K \circ Q_{K-1}$.

To now obtain our 2-frames and 2-planes, we proceed as usual.

4. $(\mathbf{v}_K, \mathbf{w}_K) = \pi(Q_K) = (Q_K \mathbf{e}_1, Q_K \mathbf{e}_2)$.
5. $P_K = \rho(\mathbf{v}_K, \mathbf{w}_K) = \langle Q_K \mathbf{e}_1, Q_K \mathbf{e}_2 \rangle$. ($\mathbf{e}_i$ is the ith canonical basis vector in R^p.)

Remarks. R1. As a concrete example of an appropriate measure μ, we take a discrete μ concentrated on the finite set of rotations $\text{supp}(\mu) = \{R_{ij}(\lambda_{ij}) | 1 \leqq i < j \leqq p\}$, where $\{\lambda_{ij} | 1 \leqq i < j \leqq p\} \cup \{1\}$ is a set of real numbers linearly independent over the integers. We simply set

$$\mu(R_{ij}(\lambda_{ij})) = \frac{2}{p^2 - p}$$

for all i, j with $1 \leqq i < j \leqq p$. We shall denote μ by $U\{R_{ij}(\lambda_{ij})\}$. By our discussion of the torus method, it is easy to see that supp (μ) generates a dense subgroup of $SO(p)$. Thus, Condition D is satisfied by μ.

R2. As long as $p \geqq 2$, Condition D guarantees that the distribution of Q_K (the position of the random walk at time K) approaches the invariant distribution on $SO(p)$. Precisely,

$$\lim_{n \to \infty} \mu^{*n} = \text{invariant measure}$$

where μ^{*n} denotes the nth convolution power of μ with itself, and the limit is understood in the sense of weak convergence [MAGL]. (Note: the invariant measure on $SO(p)$ is what is sometimes referred to as the Haar measure.)

R3. The random walk achieves its flexibility through the available choice of measures μ satisfying Condition D. By using such a measure with supp (μ) lying close to I_p, we may maintain a slow rate of change in the sequences of rotations, 2-frames, and 2-planes, and thus a high degree of continuity.

R4. The use of a measure $U\{R_{ij}(\lambda_{ij})\}$, as described in R1 above, has the following drawback. Regardless of the choice of parameters λ_{ij}, $1 \leqq i < j \leqq p$, the resulting random walk will be as unstraight as can be. Thus, the human viewer may experience disorientation in attempting to follow the resulting sequence of scatterplots. To remedy this, we hereby propose the use of the following type of measure μ.

B. *The smoother random walk.* For convenience, we first introduce the *size* of an orthogonal matrix $M \in SO(p)$ as follows:

$$\text{size}(M) = \max_{\mathbf{v} \neq \mathbf{0}} \{\text{angle}(\mathbf{v}, M\mathbf{v})\}$$

(where angle is always chosen to lie between 0° and 180°). Now pick any orthogonal matrix Q having size $(Q)=\varepsilon$ where ε is small. Also, pick a set of numbers λ_{ij}, $1 \leqq i<j \leqq p$, such that $\{\lambda_{ij} | 1 \leqq i<j \leqq p\} \cup \{1\}$ is linearly independent over the integers. Also, have the λ_{ij} satisfy

$$\delta \leqq \lambda_{ij} \leqq 2\delta, \qquad 1 \leqq i<j \leqq p$$

for some $\delta>0$ satisfying $\delta<\varepsilon (\delta \simeq \varepsilon^2$ seems to work well). Finally, we define μ to be

$$\mu(Q \circ R_{ij}(\lambda_{ij}))=\frac{2}{p^2-p}, \qquad 1 \leqq i<j \leqq p$$

on supp $(\mu)=\{Q \circ R_{ij}(\lambda_{ij}) | 1 \leqq i<j \leqq p\}$. We denote μ by $\cup\{Q \circ R_{ij}(\lambda_{ij})\}$.

It is easy to verify that this μ satisfies Condition D above and thus, as long as $p \geqq 2$, $\mu^{*n} \rightarrow$ invariant measure on $SO(p)$ as $n \rightarrow \infty$. The smaller the choice of ε, the smoother the grand tour will turn out to be.

4. Testing of grand tours. In order to assess the suitability of a grand tour for a specific application, we need to perform statistical tests on it. Two characteristics of particular concern to us are the *rapidity* with which a sequence of 2-planes becomes dense, and the asymptotic *uniformity* of the limiting distribution, if any. For each of these characteristics, there is a multitude of possible choices of how to measure them. We have chosen one test that we feel measures well the most important characteristic.

Rapidity. Here we rely on the following:

FACT. *Given two 2-planes P, $Q \in G_{2,p}$, the relative position of P and Q in R^p is described by two angles θ_1, θ_2 with $0 \leqq \theta_1 \leqq \theta_2 \leqq \pi/2$. Precisely, there exists a rotation $M \in SO(p)$ such that $M(P)=\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$ and $M(Q)=\langle \cos\theta_1 \mathbf{e}_1+\sin\theta_1 \mathbf{e}_3, \cos\theta_2 \mathbf{e}_2+\sin\theta_2 \mathbf{e}_4 \rangle$.*

The cosines of θ_1 and θ_2 are the correlations encountered in canonical correlation analysis. Thus, we use the terminology "canonical angles" for θ_1 and θ_2.

We define the distance between P and Q as the larger canonical angle: $d(P, Q)=\theta_2$. (It may happen that $\theta_1=\theta_2$.)

Now let $S=\{P_1, \cdots, Pn\}$ be a finite set of planes. Then we define the gap of S via

$$\operatorname{gap}(S)=\max_{P \in G_{2,p}} \min_{1 \leqq i \leqq n} \{d(P, P_i)\}.$$

(We are justified in using "min" and "max" rather than "inf" and "sup" since n is finite and $G_{2,p}$ is compact.) Let us define the ε-neighborhood of a plane P_0 to be

DEFINITION.

$$N_\varepsilon(P_0)=\{P \in G_{2,p} | d(P, P_0)<\varepsilon\}.$$

The number ε will be called the *radius* of $N_\varepsilon(P_0)$. In terms of this definition, it is clear that gap (S) is the radius of the largest neighborhood in $G_{2,p}$ which lies in the complement of the set S of planes:

$$\operatorname{gap}(S)=\sup\{\varepsilon>0 | \exists P_0 \in G_{2,p} \ni N_\varepsilon(P_0) \subset G_{2,p}-S\}.$$

Or expressed yet another way,

$$\operatorname{gap}(S)=\inf\left\{\varepsilon>0 \,\middle|\, G_{2,p}=\bigcup_{i=1}^{n} N_\varepsilon(P_i)\right\}.$$

We now use this last equation to establish lower bounds for $n=n(\varepsilon)$, where $n(\varepsilon)$ is the smallest number of planes needed to have gap $\leqq \varepsilon$. Namely,

$$G_{2,p}=\bigcup_{i=1}^{n} N_\varepsilon(P_i) \qquad \text{(except for a set of measure 0)}$$

and so

$$\operatorname{vol}(G_{2,p}) \leqq \sum_{i=1}^{n} \operatorname{vol}(N_\varepsilon(P_i))$$

where "vol" stands for the invariant (Haar) measure on $G_{2,p}$. By invariance, $\operatorname{vol}(N_\varepsilon(P_i))$ is independent of i, so

$$\operatorname{vol}(G_{2,p}) \leqq n \cdot \operatorname{vol}(N_\varepsilon(P_0))$$

where P_0 is, let us say, $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$.

Thus

$$n = n(\varepsilon) \geqq \frac{\operatorname{vol}(G_{2,p})}{\operatorname{vol}(N_\varepsilon(P_0))} \quad \text{or} \quad n(\varepsilon) \geqq [\operatorname{Prob}(d(P, P_0) < \varepsilon)]^{-1},$$

where $P \in G_{2,p}$ is distributed according to the invariant measure. It can be shown [Hote] that the canonical angles θ_1, θ_2 between P and P_0 have joint density function given by the following:

$$f(\theta_1, \theta_2) = \begin{cases} (p-2)(p-3)(\sin\theta_1 \cdot \sin\theta_2)^{p-4}(\sin^2\theta_2 - \sin^2\theta_1), & 0 \leqq \theta_1 \leqq \theta_2 \leqq \pi/2, \\ 0 \quad \text{otherwise.} \end{cases}$$

If $p = 3$, we have $\theta_1 = 0$ always, and θ_2 has density given by $g(\theta_2) = \sin\theta_2$, $0 \leqq \theta_2 \leqq \pi/2$, and $g(\theta_2) = 0$ otherwise.

We shall use the terminology "the 2-planes P and Q *lie within angle* Ang" to mean that the larger canonical angle $d(P, Q) = \theta_2$ is less than Ang (where $0 \leqq \text{Ang} \leqq \pi/2$).

In the tables in Table 1, obtained by Monte Carlo methods, the probability shown is the fraction of random pairs of 2-planes in Euclidean space of the given dimension which lie within angle Ang. The column labeled "No. of planes" gives a theoretical lower bound for the number of 2-planes which can be chosen in that Euclidean space so that all 2-planes lie within angle Ang of one of the chosen ones. Namely, that theoretical lower bound is the quantity: greatest integer in $1/\operatorname{Prob}(d(P, Q) < \text{Ang})$.

These tables should be thought of as a standard against which to measure the rapidity with which a sequence of planes becomes dense. In fact, if we set $N_{\text{poss}}(\varepsilon)$ = the smallest possible number of planes needed to achieve a gap of ε, and $N_{\text{gt}}(\varepsilon)$ = the smallest number of planes (in sequence), from some particular choice of grand tour, needed to achieve a gap of ε, we have

$$n(\varepsilon) \leqq N_{\text{poss}}(\varepsilon) \leqq N_{\text{gt}}(\varepsilon)$$

for all $\varepsilon > 0$. These inequalities are, with very few exceptions, actually strict ones.

Figures 1–6 display the gap as a function of the number of planes, for three types of grand tour: (1) planes picked at random, (2) planes picked by the torus method, and (3) planes picked via plain random walk on $SO(p)$. The gap was not, in fact, computed but was instead estimated via $\operatorname{gap}(N) \simeq \max_{1 \leqq i \leqq 100} \min_{1 \leqq j \leqq N} d(Q_i, P_j)$ where $\{Q_i\}$ is a fixed set of planes picked at random. Due to the vast quantity of computing time necessary, we have restricted the calculations to only two values of the dimension: $p = 4$ (using the average of 5 repetitions) and $p = 8$ (using the average of 3 repetitions).

TABLE 1

Dimension 3

Ang	Probability	No. of planes
5	.38E-02	263
10	.15E-01	66
15	.34E-01	30
20	.60E-01	17
25	.94E-01	11
30	.13	8
35	.18	6
40	.23	5
45	.29	4
50	.36	3
55	.43	3
60	.50	3
65	.58	2
70	.66	2
75	.74	2
80	.83	2
85	.91	2
90	1.0	2

Dimension 4

Ang	Probability	No. of planes
5	.19E-04	51684
10	.31E-03	3258
15	.15E-02	650
20	.48E-02	209
25	.12E-01	87
30	.23E-01	43
35	.42E-01	24
40	.69E-01	15
45	.11	10
50	.16	7
55	.22	5
60	.30	4
65	.39	3
70	.49	3
75	.61	2
80	.73	2
85	.86	2
90	1.0	2

Dimension 5

Ang	Probability	No. of planes
5	.11E-06	0.9E+07
10	.70E-05	143529
15	.78E-04	12805
20	.43E-03	2349
25	.16E-02	636
30	.45E-02	222
35	.11E-01	93
40	.23E-01	45
45	.43E-01	24
50	.75E-01	14
55	.12	9
60	.19	6
65	.27	4
70	.38	3
75	.51	2
80	.66	2
85	.83	2
90	1.0	2

Dimension 6

Ang	Probability	No. of planes
5	.67E-09	0.2E+10
10	.17E-06	0.6E+07
15	.42E-05	239048
20	.40E-04	24980
25	.23E-03	4407
30	.91E-03	1096
35	.29E-02	345
40	.77E-02	130
45	.18E-01	56
50	.37E-01	27
55	.70E-01	15
60	.12	9
65	.20	6
70	.30	4
75	.44	3
80	.61	2
85	.80	2
90	1.0	2

Dimension 7

Ang	Probability	No. of planes
5	.42E-11	0.2E+12
10	.42E-08	0.2E+09
15	.23E-06	0.4E+07
20	.39E-05	254912
25	.34E-04	29331
30	.19E-03	5185
35	.81E-03	1230
40	.27E-02	366
45	.77E-02	130
50	.19E-01	53
55	.42E-01	25
60	.82E-01	13
65	.15	7
70	.25	5
75	.38	3
80	.56	2
85	.77	2
90	1.0	2

Dimension 8

Ang	Probability	No. of planes
5	.27E-13	0.4E+14
10	.11E-09	0.9E+10
15	.13E-07	0.7E+08
20	.39E-06	0.3E+07
25	.52E-05	190626
30	.42E-04	23917
35	.23E-03	4305
40	.99E-03	1011
45	.34E-02	293
50	.99E-02	102
55	.25E-01	41
60	.56E-01	18
65	.11	9
70	.20	5
75	.34	3
80	.52	2
85	.75	2
90	1.0	2

TABLE 1 (*cont.*)

Dimension 9

Ang	Probability	No. of planes
5	.18E-15	0.6E+16
10	.29E-11	0.3E+12
15	.79E-09	0.1E+10
20	.40E-07	0.2E+08
25	.82E-06	0.1E+07
30	.92E-05	108458
35	.68E-04	14768
40	.36E-03	2755
45	.15E-02	657
50	.52E-02	192
55	.15E-01	67
60	.38E-01	27
65	.84E-01	12
70	.17	6
75	.30	4
80	.49	3
85	.73	2
90	1.0	2

Dimension 10

Ang	Probability	No. of planes
5	.12E-17	0.8E+18
10	.77E-13	0.1E+14
15	.47E-10	0.2E+11
20	.42E-08	0.2E+09
25	.13E-06	0.8E+07
30	.21E-05	486134
35	.20E-04	50072
40	.14E-03	7398
45	.69E-03	1453
50	.28E-02	359
55	.93E-02	108
60	.26E-01	39
65	.64E-01	16
70	.14	8
75	.27	4
80	.46	3
85	.71	2
90	1.0	2

Dimension 12

Ang	Probability	No. of planes
5	.58E-22	0.2E+23
10	.57E-16	0.2E+17
15	.17E-12	0.6E+13
20	.47E-10	0.2E+11
25	.34E-08	0.3E+09
30	.11E-06	0.9E+07
35	.18E-05	558417
40	.19E-04	52000
45	.14E-03	6921
50	.81E-03	1229
55	.36E-02	279
60	.13E-01	78
65	.38E-01	27
70	.97E-01	11
75	.21	5
80	.40	3
85	.67	2
90	1.0	2

Dimension 14

Ang	Probability	No. of planes
5	.28E-26	0.4E+27
10	.44E-19	0.2E+20
15	.65E-15	0 2E+16
20	.55E-12	0.2E+13
25	.93E-10	0.1E+11
30	.56E-08	0.2E+09
35	.17E-06	0.6E+07
40	.28E-05	355788
45	.31E-04	32130
50	.24E-03	4121
55	.14E-02	708
60	.64E-02	157
65	.23E-01	44
70	.68E-01	15
75	.17	6
80	.36	3
85	.64	2
90	1.0	2

Dimension 16

Ang	Probability	No. of planes
5	.14E-30	0.7E+31
10	.34E-22	0.3E+23
15	.25E-17	0.4E+18
20	.65E-14	0.2E+15
25	.26E-11	0.4E+12
30	.31E-09	0.3E+10
35	.16E-07	0.6E+08
40	.42E-06	0.2E+07
45	.68E-05	146196
50	.74E-04	13516
55	.57E-03	1767
60	.32E-02	311
65	.14E-01	71
70	.49E-01	21
75	.14	8
80	.32	4
85	.61	2
90	1.0	2

Dimension 20

Ang	Probability	No. of planes
5	.36E-39	0.3E+40
10	.22E-28	0.5E+29
15	.40E-22	0.3E+23
20	.95E-18	0.1E+19
25	.21E-14	0.5E+15
30	.95E-12	0.1E+13
35	.15E-09	0.7E+10
40	.98E-08	0.1E+09
45	.34E-06	0.3E+07
50	.71E-05	140243
55	.94E-04	10604
60	.85E-03	1180
65	.54E-02	184
70	.26E-01	39
75	.92E-01	11
80	.26	4
85	.57	2
90	1.0	2

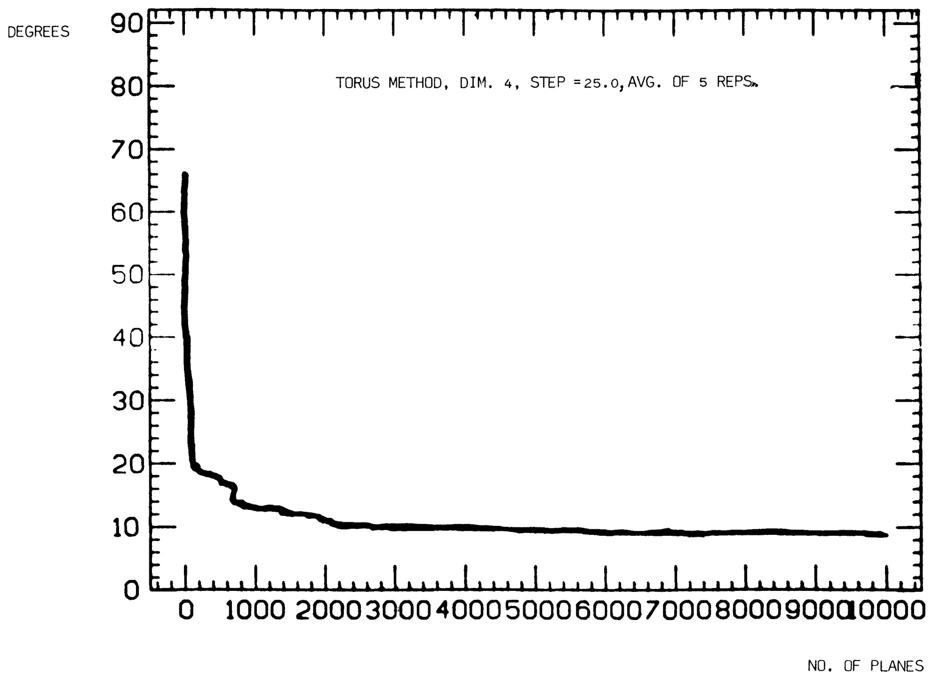
DEGREES
90
80
70
60
50
40
30
20
10
0
TORUS METHOD, DIM. 4, STEP =25.0, AVG. OF 5 REPS.
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
NO. OF PLANES

FIG. 1.

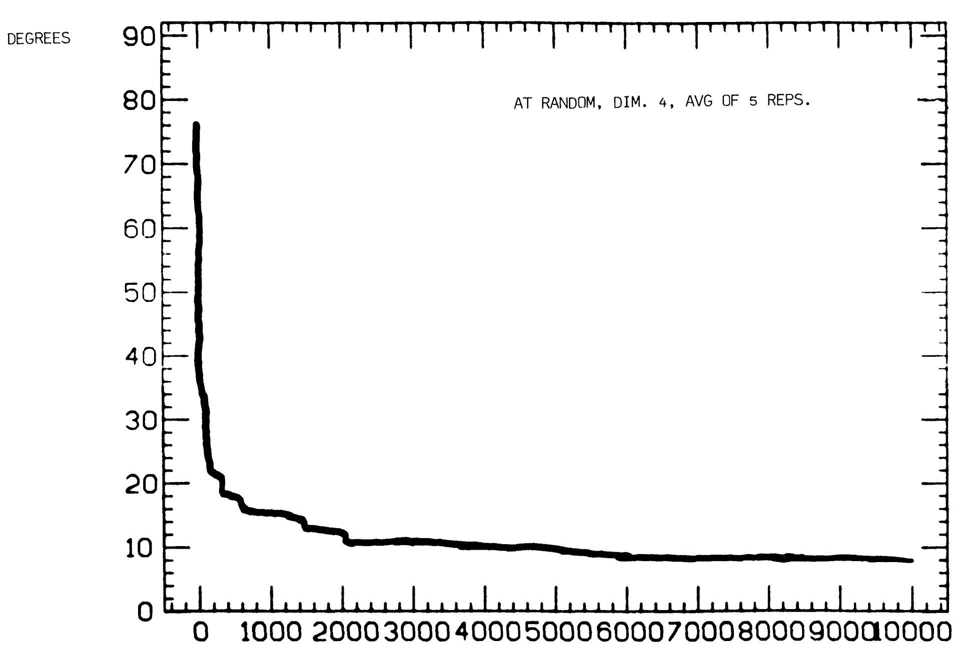
DEGREES
90
80
70
60
50
40
30
20
10
0
AT RANDOM, DIM. 4, AVG OF 5 REPS.
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
NO. OF PLANES

FIG. 2.

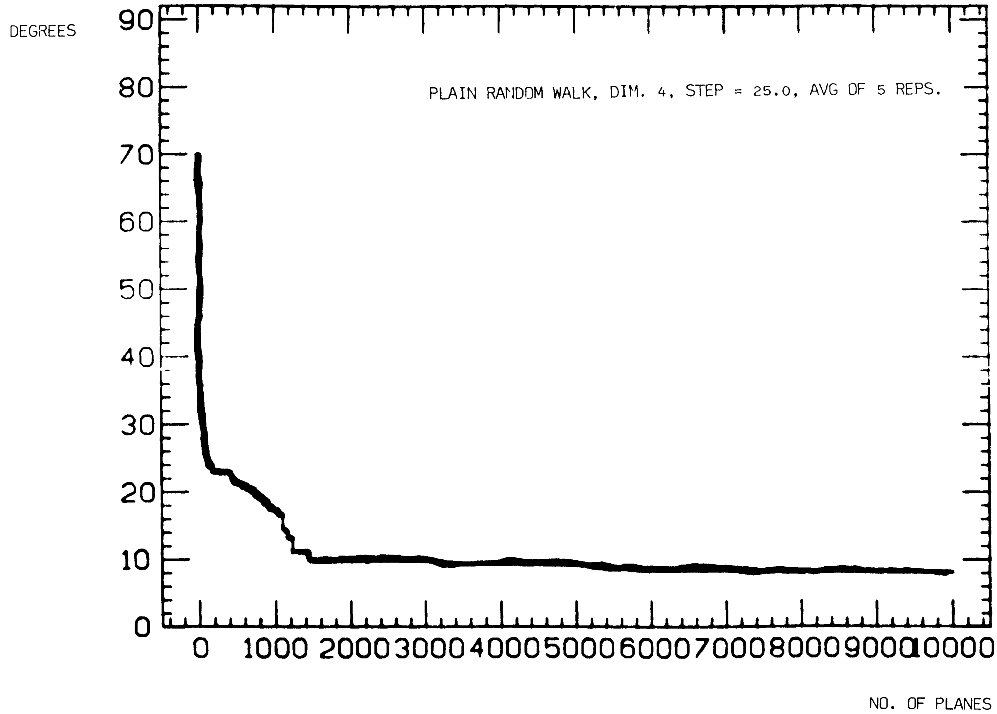
DEGREES
90
80
70
60
50
40
30
20
10
0
PLAIN RANDOM WALK, DIM. 4, STEP = 25.0, AVG OF 5 REPS.
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
NO. OF PLANES

FIG. 3.

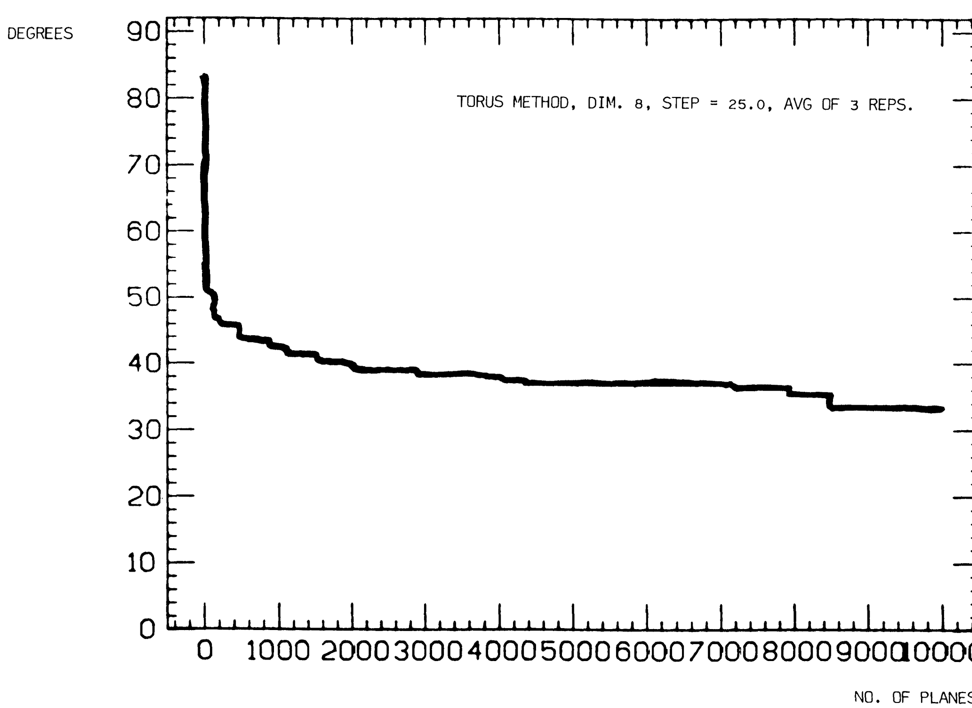
DEGREES
90
80
70
60
50
40
30
20
10
0
TORUS METHOD, DIM. 8, STEP = 25.0, AVG OF 3 REPS.
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
NO. OF PLANES

FIG. 4.

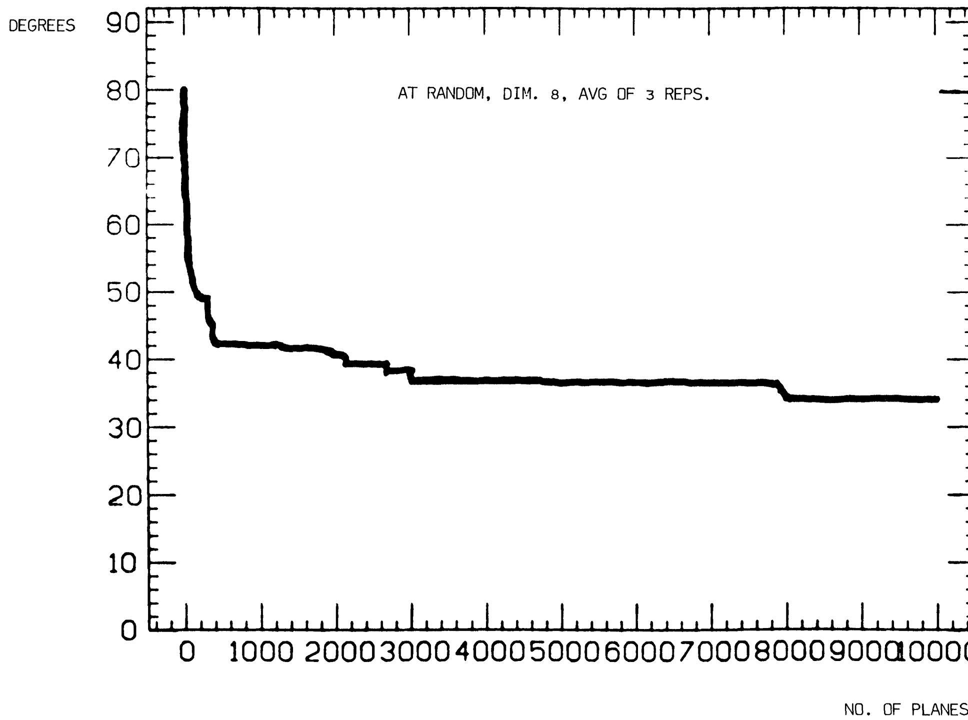
DEGREES
90
80
70
60
50
40
30
20
10
0
AT RANDOM, DIM. 8, AVG OF 3 REPS.
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
NO. OF PLANES

FIG. 5.

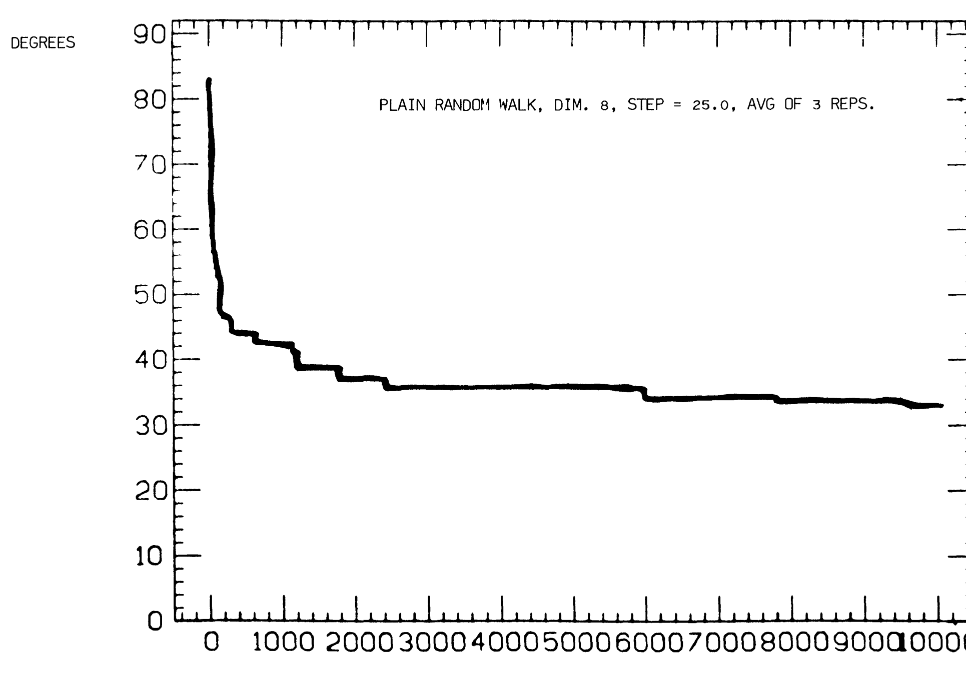
DEGREES
90
80
70
60
50
40
30
20
10
0
PLAIN RANDOM WALK, DIM. 8, STEP = 25.0, AVG OF 3 REPS.
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
NO. OF PLANES

FIG. 6.

5. Some applications of grand tours. The most basic purpose to which one may put a grand tour is to try to understand the shape of data. This understanding will presumably be applied to interpreting the data—drawing real-world conclusions.

Unfortunately, we are a long way from the point where we can do this confidently. The grand tour can be said to approximate the information content of a p-dimensional scatterplot by a time-indexed family of two-dimensional images, i.e., a movie. In order that human observers be able to interpret this kind of movie visually, a great deal of experience viewing such movies would be advantageous.

Much is still to be learned when $p=3$, and the case $p=4$ already presents a major challenge. Perhaps it would be of value to develop a *taxonomy of scatterplots* based on extensive experience with actual data. This may lead to the use of certain adjectives to describe the shapes of scatter-diagrams in greater than two dimensions. These adjectives would ideally correspond to measurements which the computer could make with great speed. An example of one such adjective-measurement pair might be the idea of "clottedness" as defined in Friedman–Tukey [PP] as their figure of merit for projection pursuit.

A useful genre of statistics may be compiled by applying a uniformly distributed grand tour to a particular scatterplot S in R^p. Let ψ be any measurement that can be applied to two-dimensional scatterplots, such as their clottedness. Then, for each 2-plane Q in R^p, we may apply ψ to the result of projecting S onto Q, obtaining $\psi(\pi_Q(S))$. As Q ranges over all 2-planes in R^p (with the invariant measure), there is a measure induced on the set of real numbers $\{\psi(\pi_Q(S))\}$. This measure carries significance especially when all coordinates represent identical units.

Statistics of this distribution of real numbers may be estimated by letting Q run through a long sequence $P_i, \cdots, P_N$ of a uniformly distributed grand tour. To take, for example, the *mean* m of this distribution, we may estimate m via

$$\hat{m} = \frac{1}{N} \sum_{i=1}^{N} \psi(\pi_i(S)),$$

where π_i denotes orthogonal projection onto the 2-plane P_i. This is a deep fact, provable by standard techniques in ergodic theory [Brei].

The advantages of using such measurements (and their corresponding adjectives) include 1) they are easy to compute, and 2) they convey an intuitive content based on the user's knowledge of two-dimensional scatterplots.

Projection pursuit methods can be described as the study of the above paradigm where the *maximum* or *minimum* of the set $\{\psi(\pi_Q(S))\}$ is the statistic of interest. These extreme values are usually sought via hill-climbing algorithms as in [PP].

One great mystery in projection pursuit is endemic to hill-climbing algorithms: how can we be confident that a local maximum is in fact the absolute maximum (or at least very near to it)? A grand tour which rapidly becomes dense in $G_{2,p}$ may be used to help with this problem. Using, e.g., the torus method with large stepsize (step $=25.0$ will work), we may let each grand tour plane $P_1, P_2, \cdots$ be the starting point for a hill-climbing procedure to maximize $\psi(\pi_Q(S))$ locally. When the local max is found, a record is kept of that value $\max_i$. One may then determine from the distribution of $\{\max_i\}$ an estimate of the absolute max.

Perhaps a better procedure would be to use the torus method with an intermediate stepsize, say step $=1.0$. Then hill-climbing may be initiated from P_i whenever

$$\psi(\pi_{i-1}(S)) < \psi(\pi_i(S)) > \psi(\pi_{i+1}(S))$$

(where π_j again denotes orthogonal projection onto P_j). Once again the local max values $\{\max_i\}$ may be stored and eventually used to estimate the absolute max.

Several films demonstrating the SLAC implementation of the grand tour have been created by the author and in collaboration with A. Buja [AsiBu].

Appendix. We describe here a method for reducing the number of matrix multiplications in the grand tour, torus method to $2p-3$.

1. Let $N=2p-3$ and think of the coordinates x_{ij} of T^N as being indexed by all pairs i, j where $i=1$ or 2, and $2 \leqq j \leqq p$ if $i=1$, but $3 \leqq j \leqq p$ if $i=2$.

2. Define a map $f: T^N \to SO(p)$ via $f(x_{1,2}, \cdots, x_{2,p}) = R_{1,2}(x_{1,2}) \circ \cdots \circ R_{2,p}(x_{2,p})$.

3. Define a map $F: T^N \to G_{2,p}$ via $F = \rho \circ \pi \circ f$ where $\pi: SO(p) \to V_{2,p}$ and $\rho: V_{2,p} \to G_{2,p}$ are as in § 3, part I above.

4. We shall prove the

THEOREM. *$F: T^N \to G_{2,p}$ is a surjection.*

Proof. Let $P \in G_{2,p}$ be an arbitrary 2-plane. We must find $x_{1,2}, \cdots, x_{2,p}$ (real numbers mod 2π) such that $R_{1,2}(x_{1,2}) \circ \cdots \circ R_{2,p}(x_{2,p}) \langle \mathbf{e}_1, \mathbf{e}_2 \rangle = P$. Letting $\theta_{i,j} = -x_{i,j}$, this is equivalent to finding $\theta_{i,j}$ (real numbers mod 2π) such that

$$R_{2,p}(\theta_{2,p}) \circ \cdots \circ R_{12}(\theta_{1,2}) P = \langle \mathbf{e}_1, \mathbf{e}_2 \rangle.$$

Now pick any orthonormal basis $\mathbf{v}$, $\mathbf{w}$ for P. First, we pick $\theta_{1,2}$ so as to satisfy

$$-\sin \theta_{1,2} v_1 + \cos \theta_{1,2} v_2 = 0.$$

This assures that $R_{1,2}(\theta_{1,2})$ annihilates the second component of v.

We similarly choose $\theta_{1,j}$ so that

$$-\sin \theta_{1,j} u_1' + \cos \theta_{1,j} v_j'$$

where v_1' = the first component of

$$R_{1,j-1}(\theta_{1,j-1}) \circ \cdots \circ R_{1,2}(\theta_{1,2}) \mathbf{v}$$

and v_j' is the jth component (in fact, $v_j' = v_j$ for $j \geqq 3$). Thus, $R_{1,p}(\theta_{1,p}) \circ \cdots \circ R_{1,2}(\theta_{1,2}) \mathbf{v}$ must have its 2nd through pth components equal to 0, and since it is a unit vector, it must, in fact, be $\mathbf{e}_1$.

We now similarly choose $\theta_{2,3}, \cdots, \theta_{2,p}$ so the 3rd through pth components of $\mathbf{w}'$ are annihilated by $R_{2,3}(\theta_{2,3}), \cdots, R_{2,p}(\theta_{2,p})$ in turn, where $\mathbf{w}'$ denotes $R_{1,p}(\theta_{1,p}) \circ \cdots \circ R_{1,2}(\theta_{1,2}) \mathbf{w}$. As a result, the vector $R_{2,p}(\theta_{2,p}) \circ \cdots \circ R_{1,2}(\theta_{1,2}) \mathbf{w}$ lies in the plane $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$. But, since the orthogonal pair $\mathbf{v}, \mathbf{w}$ is taken to an orthonormal pair by the orthogonal transformation $R_{2,p}(\theta_{2,p}) \circ \cdots \circ R_{1,2}(\theta_{1,2})$ and since $\mathbf{v}$ is taken to $\mathbf{e}_1$, we must have that $\mathbf{w}$ is taken to $\mathbf{e}_2$. Thus, we have chosen $\theta_{1,2}, \cdots, \theta_{2,p}$ so that $R_{2,p}(\theta_{2,p}) \circ \cdots \circ R_{1,2}(\theta_{1,2})$ takes the frame $\mathbf{v}, \mathbf{w}$ to the frame $\mathbf{e}_1, \mathbf{e}_2$ (more than we needed!). As a result the plane $P = \langle \mathbf{v}, \mathbf{w} \rangle$ is taken to $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$ as desired.

5. Since $F: T^N \to G_{2,p}$ is a surjection, it now follows, just as in § 3, part I, that a dense curve α in T^N will be taken by F to a dense curve $F \circ \alpha$ in $G_{2,p}$. We then just let the Kth plane of our grand tour be defined as

$$P_K = F(\alpha(K \cdot \text{STEP})), \qquad K = 1, 2 \cdots$$

for some appropriate choice of stepsize STEP.

Acknowledgments. The author would like to thank Persi Diaconis, Jerry Friedman, Peter J. Huber, Ingram Olkin, Mathis Thoma, and above all, Andreas Buja, for valuable conversations. He is also extremely grateful to Harriet Canfield for the lengthy task of typing this paper.

REFERENCES

[Andr] D. F. ANDREWS, *Plots of high-dimensional data*, Biometrics, 28 (1972), pp. 125–136.

[AsiBu] D. ASIMOV AND A. BUJA, *Finding structure in unstructured data* (a short film), Computation Research Group, SLAC, 1983.

[Brei] L. BREIMAN, *Probability*, Addison-Wesley, Reading, MA, 1968.

[Chev] C. CHEVALLEY, *Theory of Lie Groups*, Princeton Univ. Press, Princeton, NJ, 1946.

[FFT] J. FRIEDMAN, M. A. FISHERKELLER AND J. TUKEY, PRIM-9: *An interactive multidimensional data display and analysis system*, Proc. Fourth International Congress for Stereology, 1974.

[Hotel] H. HOTELLING, *Relations between two sets of variates*, Biometrika, 28 (1936), pp. 321–377.

[HWTN] G. H. HARDY AND E. M. WRIGHT, *Theory of Numbers*, Clarendon Press, Oxford, p. 381ff.

[MAGL] Y. GUIVARC'H, M. KEANE AND B. ROYNETTE, *Marches aléatoires sur les groupes de Lie*, Springer-Verlag, New York, 1977.

[MCTP] F. JAMES, *Monte Carlo Theory and Practice*, CERN, 1980, p. 38.

[MMCM] V. ARNOLD, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, 1978, pp. 148ff.

[PP] J. FRIEDMAN AND J. TUKEY, *A projection pursuit algorithm for exploratory data analysis*, IEEE Trans. Comp., C-23 (1974), pp. 881–890.

[TT77] J. TUKEY AND P. TUKEY, *Methods for direct and indirect graphic display for data sets in three and more dimensions*, Bell Laboratories, Murray Hill, NJ, 1977.