Estimating Effective Degrees of Freedom in Motor Systems

Robert H. Clewley*, John M. Guckenheimer, and Francisco J. Valero-Cuevas, *Member, IEEE*

Abstract—Studies of the degrees of freedom and "synergies" in musculoskeletal systems rely critically on algorithms to estimate the "dimension" of kinematic or neural data. Linear algorithms such as principal component analysis (PCA) are the most popular. However, many biological data (or realistic experimental data) may be better represented by nonlinear sets than linear subspaces. We evaluate the performance of PCA and compare it to two nonlinear algorithms [Isomap and our novel pointwise dimension estimation (PD-E)] using synthetic and motion capture data from a robotic arm with known kinematic dimensions, as well as motion capture data from human hands. We find that PCA can lead to more accurate dimension estimates when considering additional properties of the PCA residuals, instead of the dominant method of using a threshold of variance captured. In contrast to the single integer dimension estimates of PCA and Isomap, PD-E provides a distribution and range of estimates of fractal dimension that identify the heterogeneous geometric structure in the experimental data. A strength of the PD-E method is that it associates a *distribution* of dimensions to the data. Since there is no *a priori* reason to assume that the sets of interest have a single dimension, these distributions incorporate more information than a single summary statistic. Our preliminary findings suggest that fewer than ten DOFs are involved in some hand motion tasks. Contrary to common opinion regarding fractal dimension methods, PD-E yielded reasonable results with reasonable amounts of data. Given the complex nature of experimental and biological data, we conclude that it is necessary and feasible to complement PCA with methods that take into consideration the nonlinear properties of biological systems for a more robust estimation of their DOFs.

Index Terms—Data analysis, degrees of freedom (DOFs), dimension estimation, fractal dimension, musculoskeletal synergies.

I. INTRODUCTION

THE nervous system is involved with monitoring and controlling possibly many thousands of internal neural and muscular degrees of freedom (DOFs) as part of its motor control functions. For this reason, there is growing literature that seeks to identify and explain coordinated spatiotemporal patterns of motor activity that act on multiple DOFs at once when performing different tasks, possibly taking advantage of basic biomechanical properties of the musculoskeletal apparatus [1]–[6].

The presence of correlations between observable variables is often referred to as "synergies." From the perspective of this paper, the presence of synergies results in an effectively lower number of DOFs in observed motion, which has been called the "functional" or "effective" DOFs of the system [7]. In this case, we expect the dynamics of neuromuscular control systems to be constrained to subsets of relatively low dimension given that, by construction, the neuromuscular system needs to meet specific constraints when performing complex tasks such as locomotion or manipulation. Importantly, those subsets will be context dependent because they are a result of control actions [8] that by necessity change with task goals. In our consideration of motion capture data in a D-dimensional data space, the number of kinematic DOFs of the object is often substantially smaller than D, and the task goals constrain them to an even smaller *task space* whose dimension is smaller than the set of kinematically feasible postures. The geometry of the task spaces may be sufficiently complex that its size is better described by a distribution of dimensions than a single number.

The ability to use sensor data to objectively quantify the number of controlled skeletal DOFs during natural behaviour is central to the study of neural control of musculoskeletal redundancy. A long standing problem in this study is whether and how the nervous system fully exploits the numerous DOFs provided by the neuro-musculo-skeletal system. For example, several studies have sought to determine whether the nervous system couples the mechanical DOFs of the hand to simplify the control of hand shaping for grasp or sign language [4], [9], [10]. Other important problems are the estimation of dimension of the neural controller from electromyographic signals [1] or extracellular neural recordings from the brain [11]. Theories of motor learning also address problems of dimension estimation by proposing that the acquisition of complex tasks progresses by initially "freezing" some skeletal DOFs and gradually releasing them as the nervous systems is able to incorporate them into a motor task [12].

This paper discusses algorithmic methods that measure the dimension in state space occupied by experimentally observed dynamical behaviours. We compare the performance of three approaches to estimating the dimension of a dynamical system from sampled data: two established algorithms—principal component analysis (PCA) [13] and Isomap [14]—and a new algorithm that estimates pointwise dimension (PD-E).

PCA, linear regression, and *multidimensional scaling* [15] are linear methods that test whether a data set lies close to a linear

*R. H. Clewley is with the Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303-3083 USA (e-mail: rclewley@gsu.edu).

J. M. Guckenheimer is with the Department of Mathematics, Cornell University, Ithaca, NY 14853-4201 USA (e-mail: jmg16@cornell.edu).

F. J. Valero-Cuevas is with the Neuromuscular Biomechanics Laboratory, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853-7501 USA, and also with the Department of Biomedical Engineering and the Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA 90089-2905 USA (e-mail: fv24@cornell.edu; valero@usc.edu).

subspace, in which case the coordinates from this subspace can be used to parameterize the data. However, these methods do not determine whether the data may lie on a lower dimensional set within the subspace. Consider an example relevant to human movement: a single arm rotating relative to the body in a plane. The kinematic data set is a 1-D manifold, namely a circle in the plane, that does not lie close to a 1-D linear subspace. Linear methods select the 2-D plane in which the circular motion takes place as a minimal state space for these data (see [16] for details of this example). PCA provides useful information about the orientation of the plane containing the circle in this example, but it cannot correctly determine the dimension of this simple geometric object underlying this simple biological motion.

Isomap [14], *local linear embedding* (LLE) [17], and *Laplacian* or *Hessian eigenmaps* [18], [19] are methods that have been developed within the setting of machine learning and dimension reduction to find coordinate systems for *nonlinear* manifolds. They include procedures for discovering the dimension of data sets that lie on smooth (Riemannian) manifolds. Isomap is designed to find a set of global coordinates for this manifold via singular value decomposition of a matrix of interpoint distances of the data.

The underlying structure of a biomechanical system governing, say, locomotion or manipulation, may not be representable as motion in a smooth manifold. The structure might instead decompose as the union of submanifolds having different dimension for different phases in the gait cycle (e.g., swing versus double support) or grasp acquisition versus manipulation. Alternatively, if a biomechanical system exhibits chaotic motion (as has been suggested as far back as Bernstein [7, pp. 15-59] in the context of ubiquitous intertrial variability) musculoskeletal motion is expected to lie on a fractal set. This paper shows that PD-E aids in the process of exploring these kinds of geometric structure in data sets, and complements uses of established linear and nonlinear techniques.

Pointwise dimension is a quantity assigned to probability densities or measures that are defined on metric spaces. Like Isomap, algorithms for computing pointwise dimension are based upon analysis of the distances between pairs of data points. However, the way in which this information is used is quite different in the two methods. Algorithms for estimating the pointwise dimension of attractors of dynamical systems were developed in the 1980s and applied to many different empirical data sets [20]. Much of this work used a technique called "delay embedding" to manufacture multidimensional data from a single (1-D) time series. To use delay embedding effectively, the time step between successive observations and the number of successive observations to use in the embedding have to balanced to account for the sensitive dependence of solutions to initial conditions, the dimension of the attractor and the level of independence of observations made at each time step. There emerged an informal consensus that prohibitive amounts of data were required by the methods for accurately estimating the pointwise dimension of high-dimensional attractors [21]. This paper revisits the numerical estimation of pointwise dimension in the context of motion capture data, where the emphasis is upon estimating the dimension of sets that are already embedded in high-dimensional Euclidean spaces.

At a time when linear methods dominate the analysis of biomechanical data, we have investigated the ability of the PCA, Isomap, and PD-E methods to estimate the dimension of relevant synthetic data sets generated from a range of geometric objects and from experimental motion capture of a robot arm. We then use these methods to estimate bounds on the DOFs involved in some hand kinematics tasks. In this paper, we seek to characterize the phenomenological properties of pointwise dimension estimates for high-dimensional data sets in preparation for the future development of a rigorous mathematical theory. We consider the simple geometric objects as a means to validate PD-E on sets of known dimension that are adequately characterized by the well-established methods: we would like to be confident that PD-E could provide useful information about the dimension of any unknown data set, including ones of this simple nature.

II. METHODS

A. Dimension Estimation Algorithms

1) PCA: Assume that we have a data set of N observations in a D-dimensional Euclidean *data space* with $N >> D$. In the case of our motion capture data, k markers are placed upon an object and analysis of video recordings produces the spatial locations of these markers, yielding a data space of dimension $D = 3k$. PCA is a linear method for testing whether the data lie close to a linear subspace $U \subset \mathbb{R}^D$ whose dimension is $d < D$. The first step of PCA is to normalize the data and assemble data vectors into a $D \times N$ matrix $\mathbf{X}$. The next step is to calculate the eigenvalue decomposition $\mathbf{C} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}}$ of the $D \times D$ covariance matrix $\mathbf{C} = \mathrm{cov}(\mathbf{X}) = (\mathbf{X} - m)(\mathbf{X} - m)^{\mathrm{T}}$, where $m = \mathrm{mean}(\mathbf{X})$, $\mathbf{E}$ is a $D \times D$ matrix whose columns are the eigenvectors of $\mathbf{C}$, and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues ξ_i, ordered by decreasing magnitude. Projection onto the subspaces U_l spanned by the first l columns of $\mathbf{E}$ minimizes a residual of the original (normalized) data among projections onto l-dimensional subspaces of the data space and maximizes the variance of the projected data.

We define the cumulative sum of the first i eigenvalues as $s(i) = \xi_1 + \cdots + \xi_i$ for $i = 1, \ldots, D$, and denote its maximum value $\hat{s} \equiv s(D)$. From this, we define the fraction of variance explained up to dimension i as $\sigma(i) = s(i)/\hat{s}$, a monotonically increasing function of i. The corresponding residual (fraction of variance unexplained) is defined by the monotonically decreasing $\rho(i) = 1 - \sigma(i)$.

Estimating the dimension of the data set from PCA requires a criterion for choosing a minimal l for which the projected data is an acceptable "reduction" [22]. A frequent choice for this criterion is to set a variance capture threshold, given by the algorithmic parameter $\tau < 1$ such that $\sigma(l) > \tau$ (e.g., [4]). A second choice that is seldom used in the biomechanics literature (e.g., see [16]) is to select a value of l for which there is a "knee" (i.e., reduction in slope) in a linear-log graph of the residuals $\rho(i)$: i.e., the quantities $\rho(i) - \rho(i+1)$ are substantially larger for $i < l$ than for $i > l = 1$. This method is better tuned to the scaling properties of an individual data set. For PCA, we implement this criterion by computing the second differences of $\log(\rho(i))$ and

determine when these are larger than a threshold given by an algorithmic parameter γ. Where there are one or more consecutive second differences larger than γ, we declare there to be a knee at the local maximum of the second differences (knee positions can be determined by other algorithms [23] or by eye [24]). We found that the value $\gamma = 0.1$ caused the algorithm to select knee positions that corresponded to positions that we judged by eye.

2) Isomap: The Isomap (isometric mapping) algorithm seeks to reconstruct the Riemannian metric on a submanifold of the data space and find global coordinates that preserve this metric. One assumes that the data set of N points in $\mathbb{R}^D$ lies on a submanifold which is sampled densely enough that the Euclidean distance between near neighbours in the data set approximates distance along the manifold. Neighbourhoods consisting of these near neighbours are encoded in a "neighbourhood graph" with N vertices, one for each data point. Vertices are connected by undirected edges in this graph in one of two ways: (1) vertex v_i is connected to its K nearest neighbours in the data set, or (2) v_i is connected to vertices v_j for which the corresponding distances satisfy $\|x_i - x_j\| > \varepsilon$. Thus, the parameter of the algorithm is either K or ε. Geodesic distance between x_i and x_j is then estimated by minimizing the sum of distances along chains of points $x_i = x_0, \ldots, x_l = x_j$ which come from paths in the neighbourhood graph. The resulting distances form the matrix $\mathbf{G}(i,j)$. The neighbourhood graph may be disconnected, in which case the data is partitioned by components of the neighbourhood graph for further analysis. Here we retain only the component with the largest number of points.

Isomap then uses the classical multidimensional scaling method on the matrix $\mathbf{G}$, producing a singular value decomposition. We estimate the dimension of the data set with Isomap using a method similar to that described above for PCA, but the residual variance is defined differently. The fraction of variance captured $\sigma(l)$ for Isomap measures how much the matrix of L^2 distances between the first l singular vectors of the multidimensional scaling decomposition covaries with $\mathbf{G}$. The residual variance is then $1 - \sigma(l)$, which need not be a monotonically decreasing function of l. We search for either a minimum (when the function is nonmonotonic) or a point of maximum curvature (when the function is monotonic). We use the same criterion for detecting a knee in a linear plot of the residual variance using $\gamma = 1$.

In all our tests with the Isomap algorithm, we selected evenly-spaced "landmark" points in the data at a sample rate of 1 per ten regular data points. As recommended by Tenenbaum *et al.* [14], this results in many more landmark points than the expected dimension of the data and also many fewer than N.

Isomap can be run using either a selection of the neighbourhood radius ε or the number of nearest neighbours K as the principal parameter. As outlined by Tenenbaum *et al.* [25], we made a tradeoff between two cost functions in order to select these Isomap parameters appropriately: the fraction of the variance in geodesic distance estimates not accounted for in the Euclidean embedding, and the fraction of points not included in the largest connected component of the neighbourhood graph and, thus, not included in the Euclidean embedding of that component.

If K or ε are chosen large enough that all interpoint distances are retained, then the identity map gives the manifold metric of the sampled data and Isomap will detect only the dimension of a linear subspace containing the data. Similarly, when these parameters are chosen small enough so that only a very few interpoint distances are retained, the graph of neighbouring points becomes disconnected or the estimation of geodesic distances along a manifold are no longer accurate. When these instances arise in Section III, we will simply indicate that Isomap failed to produce a dimension estimate.

3) Pointwise Dimension Estimation: We now describe a new empirical method for estimating the pointwise dimension of a data set of N points in $\mathbb{R}^D$ which we refer to as *pointwise dimension estimation* (PD-E).

Algorithms for estimating the *pointwise* and *correlation* dimensions of data sets assign dimensions to attractors of dynamical systems [26]–[29] in the setting of *measures* or *probability densities* in a metric space, and assume that the data whose dimension is being determined is distributed like independent samples of the measure. They do not make use of the temporal structure of trajectories and can be applied to arbitrary data sets that give discrete approximations to a probability measure μ. We make the hypothesis that there is a *task space T* consisting of the closure of points in phase space that would be visited if a task were repeated an infinite number of times or performed forever. Moreover, we assume that there is a *task density μ* that describes the frequency in which different subsets of T are visited. In practice, this means that the μ-measure of a set $S \subset T$ (we call this the *volume* of S) can be approximated by the *proportion* of data points that lie in S. Methods such as PCA, Isomap, and LLE presume that, in the absence of "noise," input data represent samples from a geometric set with the structure of a Riemannian manifold. In contrast, pointwise dimension makes sense for measures supported by a much larger collection of sets. In contrast to PD-E, PCA, Isomap, and LLE do not explicitly utilize the distribution of observed points on the manifold. Nonetheless, the results of their analyses are affected by this distribution.

The pointwise dimension of $x \in T$ is defined by measuring the growth rate of balls $B_x(r)$ of radius r centered at x as a function of r. The dimension $d_\mu(x)$ of μ at x is

$$d_\mu(x) = \lim_{r \to 0} \frac{\log(\mu(B_x(r)))}{\log(r)}. \tag{1}$$

This limit may not exist and it may not be the same for all points of T. When it does exist, it reflects a power law scaling in which the volume of balls is proportional to r^d.

We adopt a pragmatic approach to defining pointwise dimension in the context of experimentally-obtained data sets T having N points. Given a reference point x, the distances between x and all other $N - 1$ points y in the data set are calculated and sorted. If r_k is the kth distance in the sorted list, then we estimate $m_x(r_k) = \mu(B_x(r_k)) = k/(N - 1)$. This is the key assumption that we make about how the time series of observed trajectories approximate the task density. A dimension estimate of T for reference point x is the asymptotic slope of $\log(\mu(B_x(r)))$ versus $\log(r)$. The scaling relationship $m_x(r) \sim r^d$ is equivalent to $\log(m_x(r)) = d\log(r) + c$ for some constant c. Thus, if there is a good linear regression fit of $\log(m_x(r_k))$ to $\log(r_k)$, then the slope of this line is
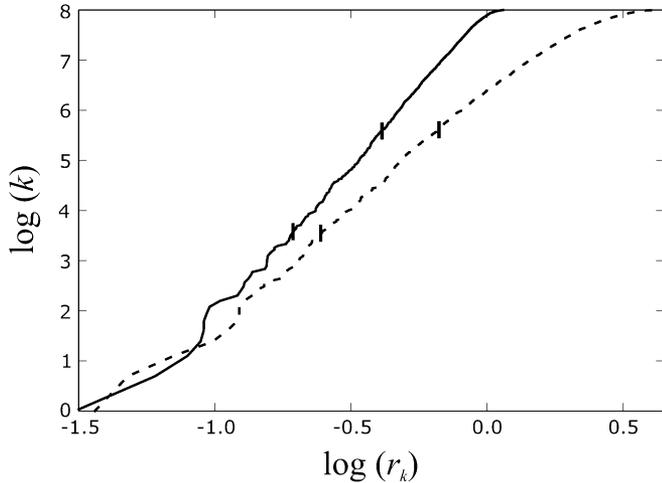
Fig. 1. Two $r - V$ curves for points randomly distributed in a 6-D ball. The solid curve has a reference point at the center of the ball, and the dotted curve has reference point $(0.57, -0.14, -0.12, 0.37, 0.66, 0.03)^{\mathrm{T}}$. The slopes of secants between the two vertical markers are 6.1 and 5.0, respectively.

taken as an estimate for $d_\mu(x)$. We call plots of $\log(k)$ versus $\log(r_k)$ $r - V$ *curves.*

If the $r - V$ curves are linear and have the same slope for all reference points, then this common slope is the pointwise dimension of the data set. However, there are inevitable statistical fluctuations and other sources of deviations of the $r - V$ curves from linear functions that occur in this procedure. Fig. 1 shows two $r - V$ curves for a data set of 3000 independent samples from the uniform probability distribution in a 6-D unit ball. One curve has its reference point at the center of the ball, whereas the second curve has a randomly chosen reference point. The curves have substantial fluctuations from a straight line at small values of $\log(r)$. The second curve deviates from a straight line also at large values of $\log(r)$. This example points to the need for additional analysis to extract good estimates of pointwise dimension from the $r - V$ curves.

Among the sources of variability in the slopes of the $r - V$ curves are the following.

1) Sampling errors that reflect the difference between the discrete data set and the probability measure μ.
2) Noise in the data yields measures that are D-dimensional, but only on scales comparable to the amplitude of the noise.
3) The pointwise dimension of the measure μ may depend on the reference point and may not exist. This happens in multifractal attractors of dynamical systems [30].
4) The "shape" of the dataset and μ affect the slope of the log-log plot at larger distances from the reference point. In Fig. 1, the growth of V slows as r reaches the distance of the reference point to the boundary of the manifold.

In the absence of firm mathematical foundations for estimating pointwise dimension, we have pursued empirical tests on observational and simulated data. We have experimented with techniques for selecting "scaling" regions of the $r - V$ curves that exclude small distances subject to large sampling fluctuations and noise, and large distances where the global shape of the object plays a dominant role in determining the relationship between volume and radius. We have also experi-

mented with ways of representing the statistical distribution of slopes with the scaling regions of $r - V$ curves. We assume that a random selection of a moderate number of reference points suffices to approximate the distribution of these slopes for the measure μ. Unless otherwise stated we select $0.2N$ reference points.

In this paper, we characterize the varying slopes of the $r - V$ curves in the following way. For the $r - V$ curve of each reference point x, we ignore the five points closest to x, and the furthest 30%. For fractal measures the $r - V$ curves may not be smooth, so we estimate their slopes with secants. Specifically, we use the slopes of secants that lie between the excluded region and have projections on the $\log(k)$ axis of length $\log(4)$. The minimum and maximum of the slopes of the secants are recorded for reference point x, and are denoted d_x^0 and d_x^1, respectively. We then calculate the minimum, maximum, mean, median, and interquartile range of the sets $(\mathrm{mins\ slopes}) = \{d_x^0\}_x$ and $(\mathrm{max\ slopes}) = \{d_x^0\}_x$ defined over the range of reference points x. We plot representative $r - V$ curves for reference points corresponding to the extrema and the means of these sets. Scatter plots of all (d_x^0, d_x^1) pairs as a function of x illustrate the distribution of slopes found. Points in the scatter plot corresponding to the extrema and means are highlighted using a color code (see Fig. 4).

We note that an alternative approach for assigning slopes to $r - V$ curves based on linear regression was also tested on the data presented here. That approach produced estimates of dimension within the range of the method described here, and is not described further for the sake of brevity.

B. Computer Generated Synthetic Data

We tested the dimension estimation algorithms on independent samples from measures of known dimension. The test measures we used are uniform distributions on 6-D rectangular solids and balls. We analyzed points uniformly distributed in a rectangular solid with sides of unit length, and from one that has four sides one fifth of the length of the remaining sides having unit length. This enables us to explore the fact that the relative length scales of different directions in the data are an issue for dimension estimation algorithms. We investigated sample sizes between 2000 and 8000 points.

C. Motion Capture Data for Robot Arm

An AdeptSix 300 robot arm with six rotational joints was used to produce motion capture data. These data sets tested our analytical techniques on a real mechanical system that is kinematically similar to a musculoskeletal limb.

The "home" configuration of the robot arm can be seen in Fig. 3(a). Fig. 2 shows a schematic diagram of the arm in this configuration, showing the local Euclidean coordinate frames defined around each link. The total length of the links is approximately 800 mm.

Three reflective markers were attached around each joint (see Fig. 3) to track the robot's posture by a 4-camera optical motion capture system manufactured by Vicon (Vcams and Vicon Workstation, Vicon Peak, Lake Forest, CA). Table I provide details of the reflective marker positions using the local coordinate axes for the joints. Marker data were captured at a rate of
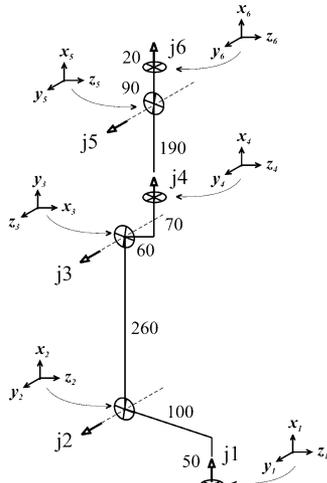
Fig. 2. Schematic diagram of the physical dimensions of the robot arm (units are in millimeters), including indication of joint axes and their local Euclidean coordinate frames (used for specifying marker positions).
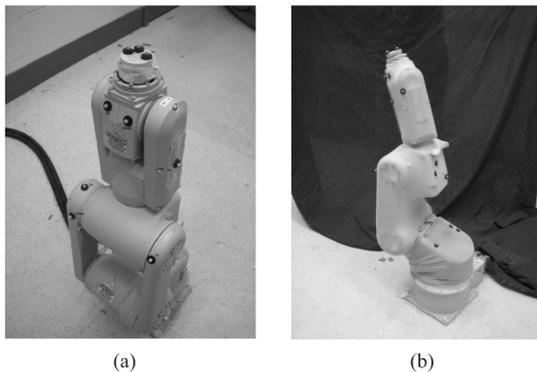

Fig. 3. (a) "Home" configuration of the AdeptSix 300 robot arm, showing some of the reflective markers used for 3-D motion capture. (b) Front view of the robot arm covered in an elastic sheath and marker placement.

TABLE I
APPROXIMATE MARKER POSITIONS RELATIVE TO JOINT AXES IN LOCAL
EUCLIDEAN JOINT COORDINATES (UNITS ARE IN MILLIMETERS)

Joint	Marker 1	Marker 2	Marker 3
1	(70, -100, 20)	(64, 100, 0)	(58, 100, -20)
2	(275, -120, 40)	(255, 110, 45)	(300, 110, 5)
3	(-120, 30, -55)	(-60, -30, -55)	(30, 15, -50)
4	(100, 75, 50)	(70, 75, 25)	(70, -85, -20)
5	(10, -30, 65)	(20, 30, 60)	(10, 30, -55)
6	(30, -35, 0)	(30, -20, -30)	(30, 32, 7)

100 Hz. Only the frames in which all markers were visible and properly reconstructed were kept in the final data set. The mean calibration residual of the Vicon marker reconstruction is less than 0.2 mm. The standard deviation in the reconstructed distance between two markers on a rigid object is approximately 0.05 mm.

Two experiments were performed with the AdeptSix 300. In the first, the robot arm was programmed to move to a succession of joint angles in a random walk that cyclically varies

a single joint angle at each step. Joints were constrained to the ranges $\pm 30°, \pm 15°, \pm 15°, \pm 45°, \pm 15°, \pm 120°$, respectively. The transition time between target postures was approximately 1/3 s. Robot movement did not exhibit extraneous oscillations.

The joint angle targets chosen in the first experiment were recorded so that the same sequence of targets could be reproduced in a second experiment. The second experiment differed in that we put the entire robot arm inside a tight elastic sheath (white hosiery made by L'eggs, approximately 80% nylon and 20% spandex, attached by elastic bands around joints 2 and 4), and re-attached the markers in positions as close as feasible to their prior positions in the nominal configuration (see Fig. 3). The sheath was intended to provide a source of systematic noise in the measurement of the robot's actual motion, in this case to mimic the effects of skin in the reconstruction of animal skeletal motion using surface-mounted markers.

Two and a half hours of data were collected from each experiment, but this was resampled at a rate of approximately one frame per 3 s, resulting in a data set of approximately 4000 points. We did not filter the kinematic data.

D. Virtual Robot Motion

We further tested our methods with synthetic data of an ideal simulated robot arm without an elastic sheath. We reconstructed the geometry of the AdeptSix 300 robot in a kinematic chain model of the joints, placing the same number of markers in approximately the same positions. The model was positioned by setting the six joint angles, and the forward kinematic transformation from angle space into Euclidean marker space was performed to generate marker positions of the virtual robot.

We used two methods to generate joint angles of the arm. One method was the same random walk protocol used for the physical robot. The second was to use independent samples of a uniform measure in joint space. The same limits on the joint angles were used as for the physical robot. The chosen angles were mapped into marker space to obtain the data set.

E. Constrained Hand Motion

We used the Vicon 3-D motion capture system to record kinematic time-series data of two informed and consenting subjects asked to perform three tasks. The protocol was approved by Cornell University's University Committee on Human Subjects. The subjects held their wrists in a fixed position while moving their fingers. Five reflective markers were placed on each finger (one at the fingertip and two between each of the joints), three on the thumb, and four additional markers were placed on the back of the hand (a total of 27 markers).

The first task was simultaneous "random" movement of fingers close to the plane of the palm. The other two tasks were the simulation of typing on a computer keyboard and the simulation of manipulation of a track ball. These tasks were performed for approximately 20 min in four 5-min segments, and the resulting data sets combined and resampled to select three frames per second. We discarded frames in which hand postures caused a failure in the reconstruction of any marker's position. This process resulted in final data sets containing approximately 8000 points.
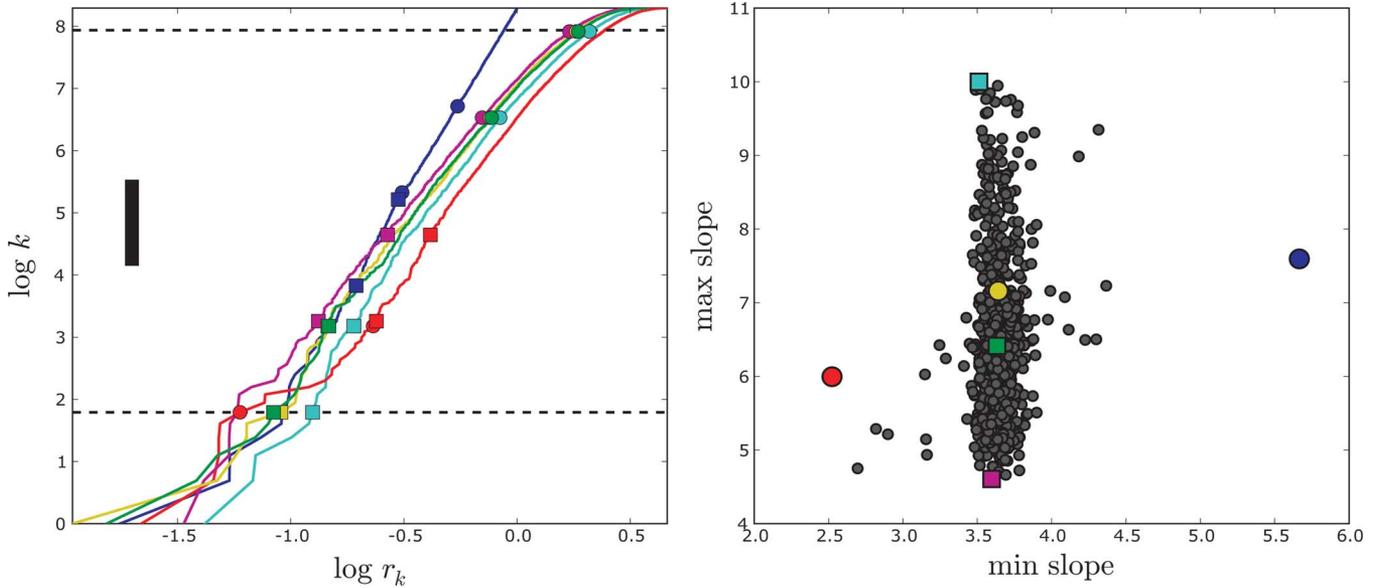
Fig. 4. PD-E analysis of sample points from a 6-D unit ball. The left panel shows $r - V$ curves for the color-coded points in the scatter plot in the right panel. These highlighted points indicate the minimum, maximum and mean of the minimum and maximum slopes. Round (square) markers indicate statistics relating to the minimum (maximum) slopes. The color coding is as follows: red = minimum of (min slopes); yellow = mean of (min slopes); blue = max of (min slopes); magenta = min of (max slopes); green = mean of (max slopes); cyan = max of (max slopes). The secant end points estimating the minimum and maximum slopes of the $r - V$ curves are indicated by round and square markers, resp. The solid bar indicates the vertical extent of the secants, equal to $\log(4)$. Dotted lines mark the closest and furthest nearest neighbours considered in the estimation of slopes. The broad distribution of maximum slopes in the scatter plot can be attributed mostly to the distance of the associated reference point x from the center of the ball.
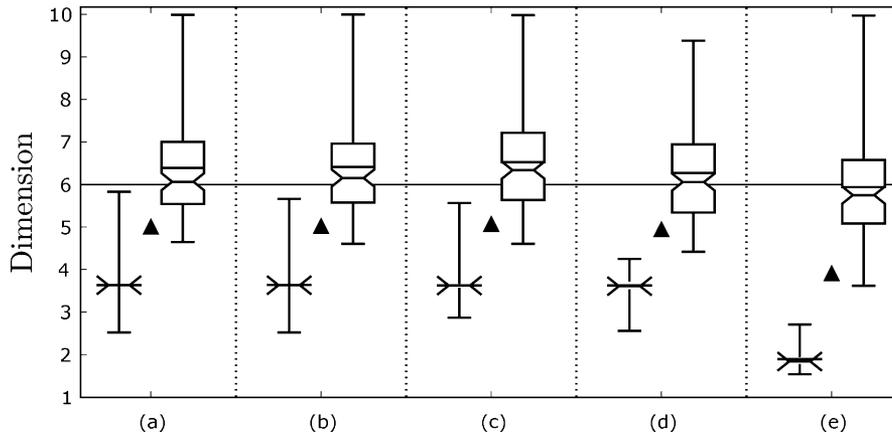

Fig. 5. PD-E dimension estimates for sample points from 6-D solids: (a) a unit ball with $N = 8000$, (b) a unit ball with $N = 4000$, (c) a unit ball with $N = 2000$, (d) a unit cube with $N = 2000$, and (e) a rectangular solid having two sides of unit length and four sides of length 0.2, with $N = 2000$. For each solid, two box-and-whisker plots are shown. The left plot indicates the distribution of (min slopes), the right indicates that of (min slopes). In each case, the whiskers mark the extent of the data (from minimum value to maximum), the boxes mark the interquartile range. The notches indicate the median value. The horizontal line in the box indicates the mean value. When the interquartile range is very small, the top and bottom of the box is not drawn for the sake of clarity: the range is still apparent by the distance from the median to the beginning of each whisker. The black triangle between each pair of box plots indicates the mean of the data from (min slopes) and (min slopes) taken together. The thin horizontal line across each panel indicates the known dimension D of the data set. Panels (a)–(c) show the relative insensitivity of PD-E results on the number of data points. Panel (d) shows that a nonsmooth boundary in the data does not significantly affect the results, compared to the smooth boundary in (c). Panel (e) shows that a high ratio of side lengths distorts the distribution of PD-E dimension estimates.

III. RESULTS

A. Computer Generated Synthetic Data

We expect that Isomap and PCA would not detect that dimension reduction is appropriate for data sampled randomly from a uniform probability distribution on a ball or rectangular solid in $\mathbb{R}^D$. We tested data sets of N independent random samples from a 6-D unit ball, using $N = 2000, 4000$, and 8000. PCA at the 90% variance capture threshold determined

$d = 6$, and graphs of PCA residuals and Isomap residual variances indicated no "knees." Thus, using PCA or Isomap in this manner predicts that dimension reduction is not appropriate. However, PCA at the 80% variance capture threshold determined $d = 5$, implying that for a sufficiently low threshold this use of PCA incorrectly predicts that dimension reduction is appropriate. PD-E analysis accurately estimated d by the median of the maximum slopes of $r - V$ curves (summarized in Figs. 4 and 5).
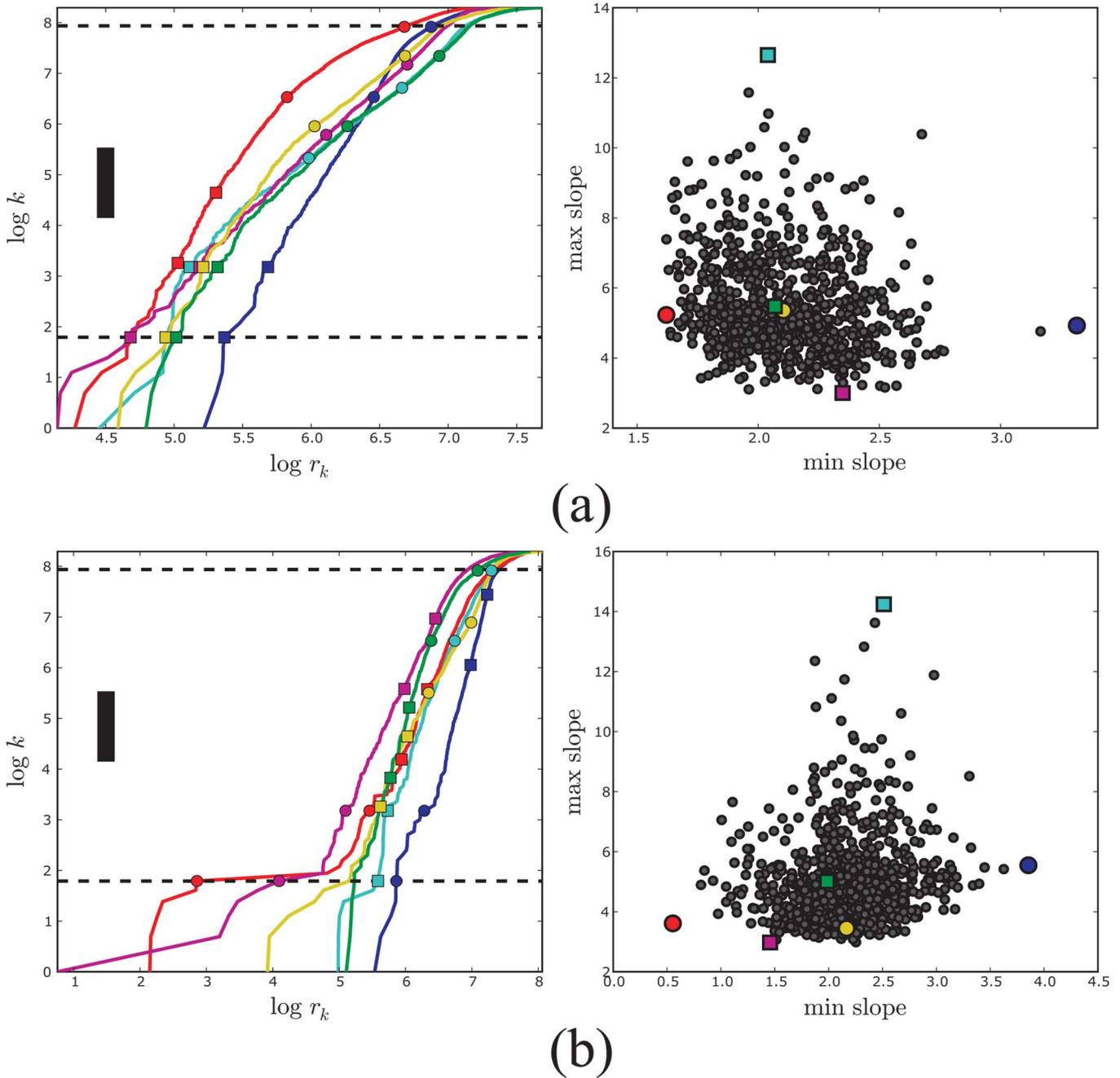
Fig. 6. PD-E analysis for the virtual robot. (a) Uniform distribution of joint angles (Test 1). (b) Joint angles determined by a random walk in joint angle space, and almost flat regions of some $r - V$ curves are observed for small radii.

B. Robot Arm Data

In this section, we analyze data sets for three motion tasks, each consisting of 4000 points. Representative $r - V$ curves and scatter plots from PD-E analysis of the data are shown in Fig. 6, and the corresponding analysis is presented in Fig. 7.

In Test 1, we consider data from a virtual robot that is randomly sampled from a uniform distribution in joint space. In Test 2, we slaved the position of joint 2 to be a smooth function of the position of joint 3, according to $\theta_2 = \theta_3^3$. In this test, the number of DOFs of the system are reduced by one, and PD-E, Isomap, and PCA (using residual variances) all successfully detected this.

In Test 3, we generated data for the virtual robot arm by performing a random walk in joint angle space. One major difference in these results is the presence of $r-V$ curves with very flat regions [Fig. 6(b)]. These regions cause the distribution of minimum slopes d_x^0 to include values near zero, and as these regions contain so few points their slopes are not interpreted as dimension estimates. The initially steep rise of $\log(k)$ as a function of r_k before a flat region suggests that there is a small cluster of data points that are closely spaced in comparison with the typical distances between other points in the data set. The centers of the dense clusters are located at distances comparable to the mean interpoint distances between all points in the data set. This
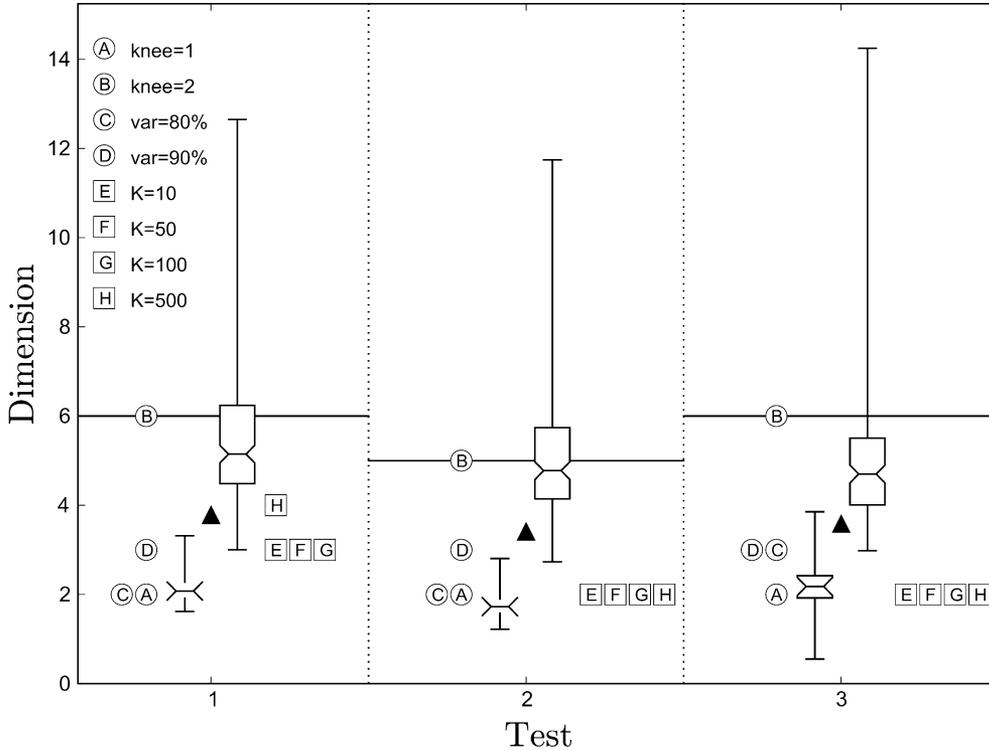
Fig. 7. Dimension estimates of virtual robot data sets. Test 1 uses a uniform distribution of joint angles. In Test 2, the position of joint 2 is slaved to be a smooth function of the position of joint 3. In Test 3, joint angles are determined by a random walk in joint space. The circular (square) markers indicate the dimension estimation results for PCA (Isomap).

suggests that flat regions in these $r - V$ curves may be due to the fractal structure of random walks and the low dimension of Brownian sample paths [31].

Fig. 8 shows the results of our analysis on the experimental motion capture data from the AdeptSix300 robot arm. The mean values of the maximum and minimum slopes are close to those predicted from the virtual robot tests, and the qualitative pattern of point distribution in the scatter plots is similar to that in the virtual robot results when the method of joint angle generation is the same [summarized in Fig. 6(b)]. The addition of the elastic sheath to the robot did not significantly change any of the dimension estimates, and made almost no difference to the interquartile ranges of (min slopes) or (max slopes).

C. Hand Motion

Our analysis of four data sets for human hand motion using $N = 4000$ is given in Figs. 9–11. This suggests that the dimension of hand motion is less than 11, and probably around 6. The histograms of d_x^0 and d_x^1 also show that there is substantial variance in the distributions of these slopes between tasks. Future work will investigate this as a possible indication that the appropriate dimension reduction for hand motion may be task dependent.

In the scatter plots, we observe a large number of points for which min (min slopes) ≈ 0.5. The red $r - V$ curves contain the minimum d_x^0 secant slopes, which occur in the flat regions of those curves. In the study of the synthetic data sets, we identified that such flat regions correspond to localized subsets of a small number of closely-spaced points, and not an indication of low dimension. In this case, there appear to be many such small

subsets. It will be instructive in future work to identify which hand postures are associated with these subsets.

IV. DISCUSSION

The characterization of the number of effective degrees of freedom (DOFs) of a limb or anatomical system under neuro\-muscular control is a necessary prelude to understanding neuromuscular control and developing dynamical models for the motion of a biomechanical system. The measurement of these active DOFs is complicated by the fact that voluntary function limits a system to a subset of its kinematically possible states which themselves form a nonlinear space determined by mechanical constraints. To complement the estimation of the effective DOFs of a system using linear methods such as PCA, we compared two ways of using PCA with two nonlinear methods [Isomap and the novel pointwise dimension estimation (PD-E) algorithm] that are also designed for this purpose. We compared the dimension predictions of the three methods using simulated kinematic data, motion capture data from a robot arm whose DOFs are known, and for finger motions whose DOFs are unknown. The dimension estimates for the robotic data are all consistent with the known number of DOFs of these systems. The estimated dimension of the finger motions is not much larger to that obtained from the robot. This gives us confidence that the methods produce reasonable estimates for the number of effective DOFs of finger motion tasks. We shall pursue further use of the methods to compare the effective number of DOFs of different tasks and to detect the heterogeneity of their task space geometries.
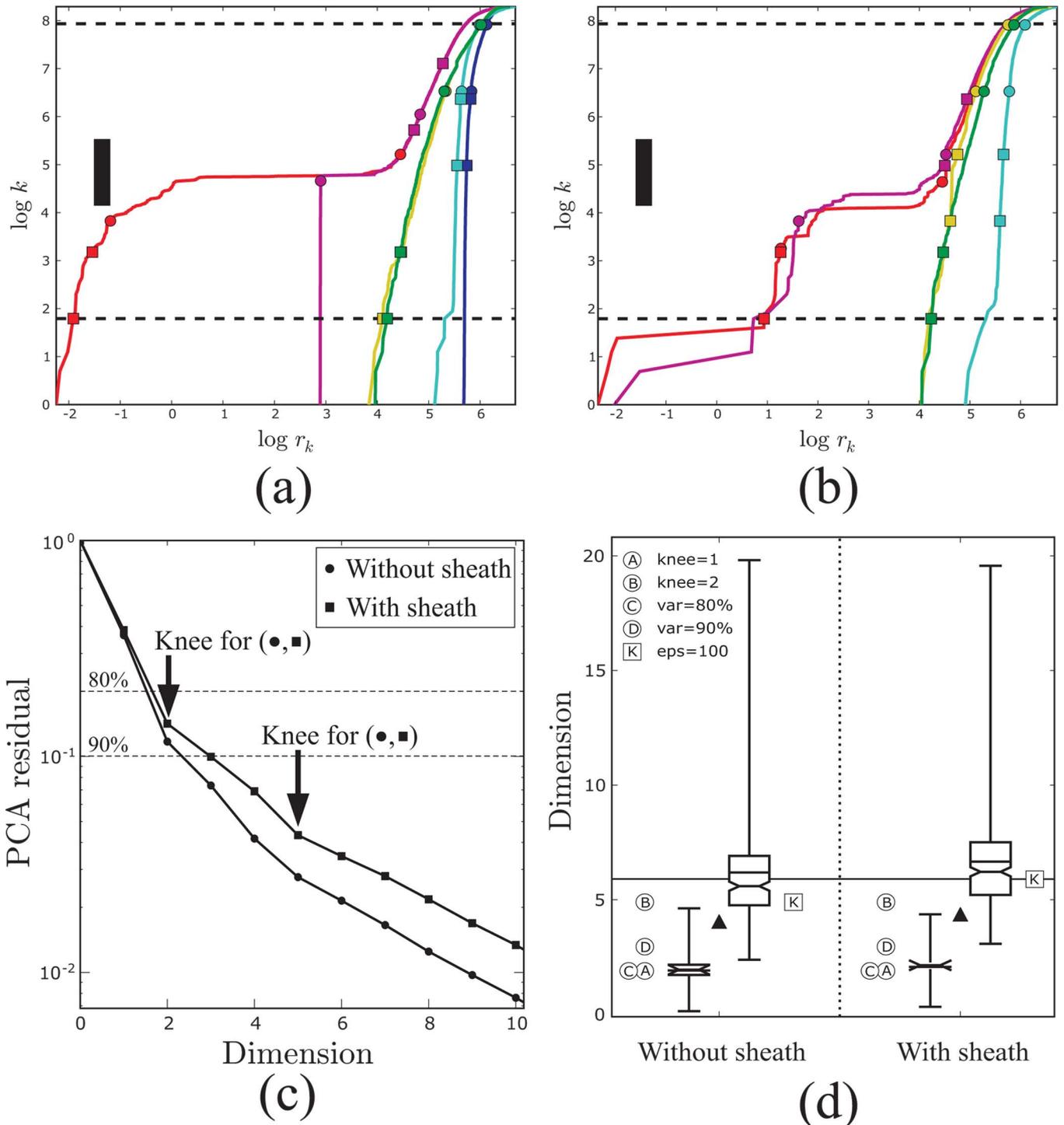
Fig. 8.   PD-E analysis of data for (a) the robot arm without an elastic sheath and (b) the robot arm covered with the sheath. (c) Residuals of d-D PCA embeddings of the robot motion data. (d) Summary of dimension estimation results. The circular (square) markers indicate the dimension estimation results for PCA (Isomap).

Using PD-E we placed a conservative upper bound on the dimension of constrained hand motions at roughly 10 for tasks involving relatively open hand postures. However, we observed a high frequency of slopes in the range of 2–3 and 5–10 in the $r - V$ curves, estimates of 1–2 and 4–7 from PCA when detecting knees in the residuals, and estimates mostly in the range of 4–7 from Isomap. Thus, the combined estimates from the three methods we have focused on suggest that the neuromus-

cular control of hand motion involves a similar number of DOFs to that estimated by Santello *et al.* [4]. In that study, the DOFs of hand motion were estimated in the very different context of a task consisting of a wide range of static hand postures for grasping objects using two methods: PCA yielded an estimate of approximately 3 DOFs, and a combination of discriminant analysis and information theory estimated an upper bound of 5 or 6 DOFs.
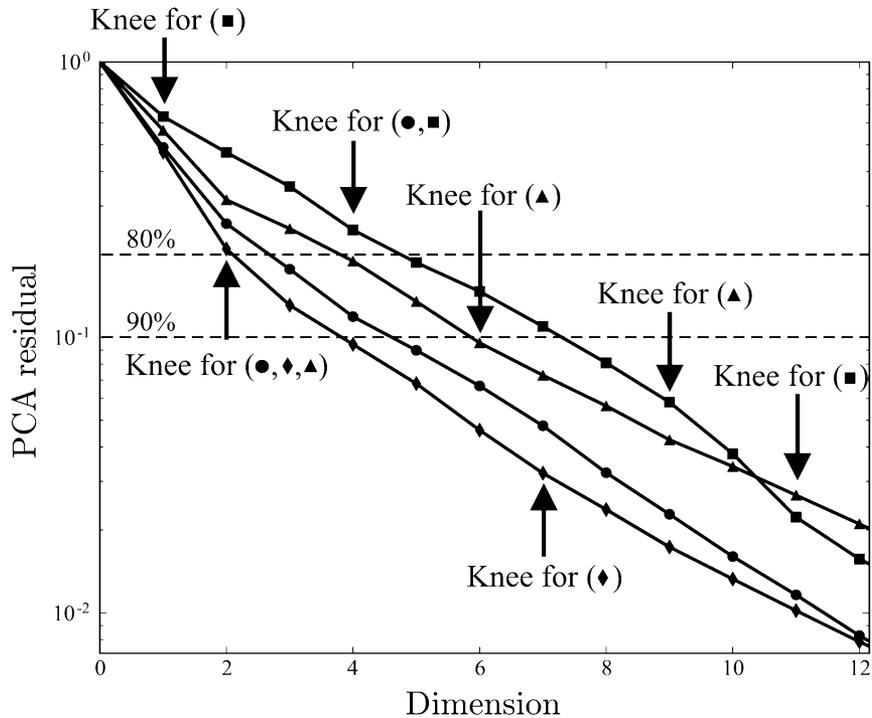
Fig. 9. PCA residuals for all hand motion tasks, showing first two knees for each task. Random motion task (subject S2): circle markers; (subject S1): square markers. Keyboard task (subject S2): triangle markers. Trackball task (subject S1): diamond markers.

The PD-E analysis discovered an interesting property of the robot data. The robot was programmed to perform a random walk in a rectangle volume of its state space. Sample paths of random walks have dimension two in unbounded state spaces of dimension larger than one. The time scale on which the random walk appears to "fill" a bounded region depends upon its volume and can be very long. Some of the PD-E $r - V$ curves from the robot display prominent flat portions that we believe are caused by the propensity of the random walk to form "clusters" of points that maintain close distances to each other, separated by "flights" between clusters. Flat portions of an $r - V$ curve represent a range of distances from its reference point that contain few points from the data set. We tested this hypothesis with our virtual robot by comparing PD-E analysis of time series that reproduced the random walk driving the robot with time series generated from independent random samples of a uniform distribution in the robot joint space. The time series of independent random samples did not produce flat portions of its $r - V$ curves. Note that the $r - V$ curves of the finger motion data do display these flat regions. This analysis bears upon the question of how long a system must be observed for the resulting trajectory to produce a good approximation to the system's task density.

There are three types of processes yielding trajectories with different time scales for this approximation. Independent random samples produce a good approximation that is independent of the geometry of the task space. If the task space is a chaotic attractor of a dynamical system, then there is a characteristic "mixing" time associated with the Lyapunov exponents for trajectories with nearby initial conditions to become statistically independent of each other [20]. If the sampling interval of the trajectory is larger than the mixing time, then the resulting time series will have the characteristics of independent random samples. In the case of a random walk, the relevant characteristic time will depend upon the square of the radius of the task space because the expected distance that a point moves in the task space in time t is proportional to $t^{1/2}$. In one of the finger motion tasks that we observed, subjects were instructed to move their fingers randomly. We suggest that the flat regions of the $r - V$ curves produced from this data reflect that their movements resembled a random walk, sampled at intervals short enough to maintain correlations between successive samples. An alternative hypothesis is that the subjects had certain "preferred" postures that have large weight in their task space because they returned to these postures repeatedly during the experiment and rested there long enough to achieve many observations at those locations.

Our implementation of the PD-E algorithm is suitable for data sets of a few thousand points, which are feasible and commonplace in biomechanical experiments. In [32], we challenged the idea that a method such as PD-E, that is based on calculations of a type of fractal dimension, requires unreasonably large amounts of data. The algorithm can analyze larger data sets than similar algorithms that calculate a complete matrix of interpoint distances for points in a data set. Nevertheless, experimental trials that are as long as possible are still necessary (for any method) to ensure that the sampling of postures from the robot or hand a) approximate the full distribution of postures assumed in the specified task; and b) are at a sufficiently low sampling rate to prevent close correlations between successive samples. As discussed above, the sampling intervals required to satisfy b) may depend qualitatively on the type of task being performed.
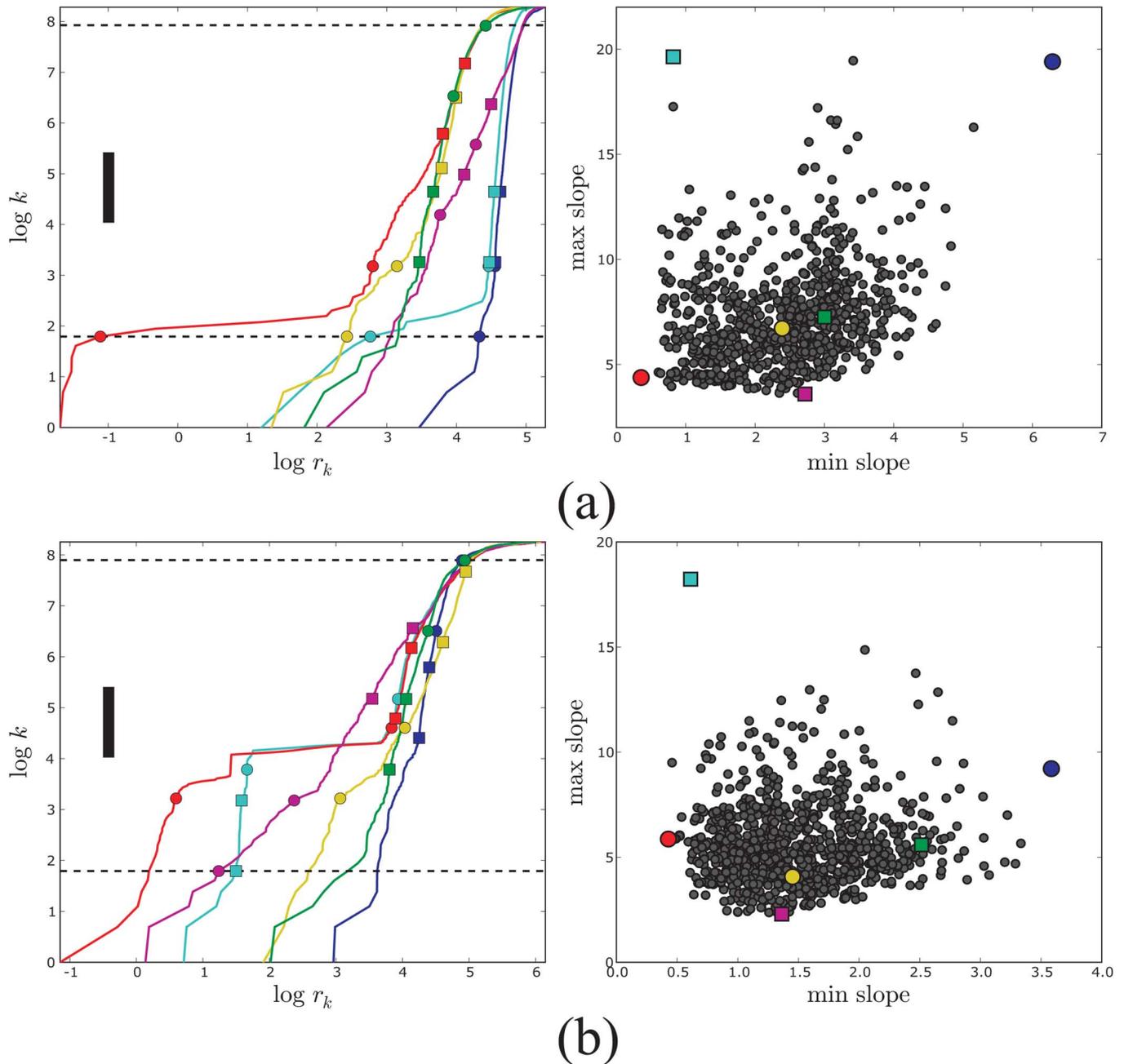
Fig. 10. PD-E analysis of data for two hand motion tasks. (a) Random motion (subject S1). (b) Trackball simulation (subject S1). Notice the presence of flat regions in some of the $r - V$ curves in both tasks, and the mild flattening of the curves for radii $\log(r) > 5$.

Estimating the effective DOFs from experimental measurements is complicated by the effects of instrumentation noise, and more importantly by "passive" DOFs of the system that are inherent in its mechanical properties. For instance, the elasticity of the skin surface on which we place reflective markers may add passive DOFs. To test the sensitivity of dimension estimation methods to this type of noise, we explore the effects of a skin-like barrier covering our robot arm upon the analysis. We did not filter our kinematic data so that we retained any synergies in the form of correlated high frequency components of motion, along with actual noise. Increased noise introduced by the skin-like barrier was apparent in the data, but had little effect on the estimated dimensions.

Our results demonstrate that the dimension of task spaces of nonlinear biological systems need not be uniquely defined. The entire distributions of interpoint distances derived from an experimentally observed data set contain valuable information. The relative simplicity of generating single-valued dimension estimates by any linear or nonlinear method belies the fact that, taken alone, those estimates may conceal important information about the finer structure of those data. For example, the task space of the system may decompose as the union of submanifolds having different dimension for different phases in the gait cycle (e.g., swing versus double support) or grasp acquisition versus manipulation. The PD-E method we have introduced produces a distribution of pointwise dimensions that reflect both
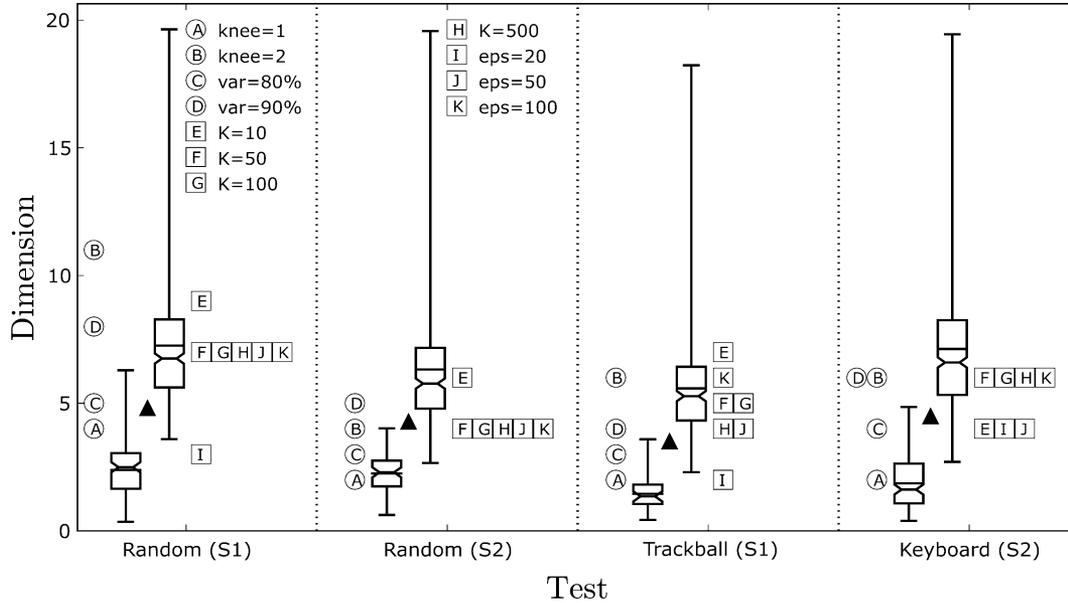
Fig. 11. Dimension estimates of hand data sets. Tasks: Random motion, simulation of trackball manipulation, simulation of computer keyboard use. The circular (square) markers indicate the dimension estimation results for PCA (Isomap).

noisy data and the heterogeneous geometry of the task spaces we observe. The complexity of the observed system will affect how much information is contained in the distribution of $r - V$ curves beyond what can be inferred from the mean, variance, maxima and minima of the slopes.

There is little rigorous theory for the effects of finite sample size in estimating dimension, even for uniform measures on smooth manifolds. Typically, the slopes of the $r - V$ curves are smaller than the pointwise dimension of the underlying measure due to "boundary effects." These effects are seen most clearly in the flattening of slopes at large radii, and is inevitable because reference points generally do not lie close to the center of a distribution even when that distribution is uniformly sampled from a symmetric geometric object such as a ball. In the tests on low-dimensional data sets of known dimension, the median value of (max slopes) gave a good estimate of the dimension. Isomap performed well on our tests with low-dimensional data, when we were able to find control parameter choices for K or ε that enabled the algorithm to converge. We compared two methods for obtaining dimension estimates using PCA. One is based on a fixed variance capture threshold τ. The second involves plotting the PCA residuals $\rho(l)$, where we looked for gaps in the spectrum of the PCA decomposition by finding "knees" in the graph. In our tests on data sets of known dimension, a dimension estimate based on the position of either the first or second knee was typically more accurate than estimates based on predetermined values of τ. However, either use of PCA overestimated the dimension of data that has nonlinear structure, for instance data sampled from a cyclic motion retracing a closed curve in state space, or from a nonsmooth manifold [32]. PD-E and Isomap performed well on tests involving cyclic motion, but only PD-E appears to be sensitive to data sets that can be decomposed into subsets having different dimensions. This structure appears as clustering in the scatter plots and a multimodal distribution of slopes in the $r - V$ curves, and in

general requires careful analysis. However, in benchmark tests, our use of minimum and maximum slope statistics was sufficient to detect the bi-modal structure of data sets decomposing to two subsets of differing dimension [32]. Tests involving measures of high dimension are also discussed at greater length in [32].

We conclude from these tests that an algorithm such as the PD-E method presented here is effective in the *qualitative* exploration of complex geometric structure in motion capture data. This method is intended to complement other methods and not to replace them. The correspondence between the estimated DOFs of the virtual and real robots suggest that we can expect improvement in the accuracy of dimension estimation for experimental data sets by continuing to refine our methods using abstract test data. Additionally, these numerical explorations will guide the development of a more systematic analysis method for the application of the pointwise dimension to high-dimensional data sets, as well as provide a focus for a rigorous mathematical theory. For instance, one development would be to follow a pointwise dimension analysis of a data set with a parameterization of the low-dimensional manifolds it identifies. This could be done using PCA or more general techniques such as kernel PCA [33], and would elucidate finer detail in the geometric structure of the manifolds and their orientation in the whole data set.

In summary, we provide several important insights into the use of fractal dimension estimates for high-dimensional data sets. We have shown that the use of pointwise dimension makes the practical analysis of such data computationally feasible, in that it requires only a few thousand data points and a modest number of reference points. In contrast to PCA, an inherently nonlinear dimension estimation method is desirable in situations where we cannot expect the dynamics generating the data to be linear. In the context of data generated by complex processes such as biomechanical systems it is preferable not to presume that attractors for the underlying nonlinear dynamical sys-

tems have the structure of a single smooth manifold. In this regard, PD-E is distinct from other nonlinear dimension estimation methods such as Isomap in that it can be applied to such nonsmooth sets, and provides information relevant to the appropriate partitioning (or "clustering") of such data sets. Last, our method does not presume that the relevant scales inherent in the data sets are known. Such scales are used as parameters by methods such Isomap, local linear embedding, etc. However, we have found that the performance of these methods can be sensitive to these parameters in a way that makes their determination difficult for the kinds of experimental data sets we consider here. Instead, these scales emerge from a PD-E analysis, and provide reasonable values for Isomap's parameters [32].

REFERENCES

[1] A. d'Avella and E. Bizzi, "Shared and specific muscle synergies in natural motor behaviors," *Proc. Nat. Acad. Sci. USA*, vol. 102, pp. 3076–3081, 2005.

[2] F. Gandolfo, F. A. Mussa-Ivaldi, and E. Bizzi, "Motor learning by field approximation," *Proc. Nat. Acad. Sci.*, vol. 93, no. 9, pp. 3843–3846, 1996.

[3] M. L. Latash, J. P. Scholz, and G. Schöner, "Motor control strategies revealed in the structure of motor variability," *Exerc. Sport Sci. Rev.*, vol. 30, no. 1, pp. 26–31, 2002.

[4] M. Santello, M. Flanders, and J. F. Soechting, "Postural hand synergies for tool use," *J. Neurosci.*, vol. 18, no. 23, pp. 10105–10115, 1998.

[5] E. Todorov and M. I. Jordan, "Optimal feedback control as a theory of motor coordination," *Nature Neurosci.*, vol. 5, pp. 1226–1235, 2002.

[6] M. C. Tresch, V. C. K. Cheung, and A. d'Avella, "Matrix factorization algorithms for the identification of muscle synergies: Evaluation on simulated and experimental data sets," *J. Neurophysiol.*, vol. 95, pp. 2199–2212, 2006.

[7] N. A. Bernstein, *The Co-Ordination and Regulation of Movements*. New York: Pergamon, 1967.

[8] E. Todorov, "Optimality principles in sensorimotor control," *Nature Neurosci.*, vol. 7, no. 9, pp. 907–915, 2004.

[9] E. Todorov and Z. Ghahramani, "Analysis of the synergies underlying complex hand manipulation," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2004, vol. 6, pp. 4637–4640.

[10] E. J. Weiss and M. Flanders, "Muscular and postural synergies of the human hand," *J. Neurophysiol.*, vol. 92, pp. 523–535, 2004.

[11] A. B. Schwartz, D. M. Taylor, and S. I. Tillery, "Extraction algorithms for cortical control of arm prosthetics," *Curr. Opin. Neurobiol.*, vol. 11, no. 6, pp. 701–707, 2001.

[12] K. M. Newell and D. E. Vaillancourt, "Dimensional change in motor learning," *Human Movement Sci.*, vol. 20, pp. 695–715, 2001.

[13] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.

[14] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[15] T. Cox and M. Cox, *Multidimensional scaling*. London, U.K.: Chapman & Hall, 1994.

[16] A. Daffertshofer, C. Lamoth, O. Meijer, and P. Beek, "PCA in studying coordination and variability: A tutorial," *Clin. Biomech.*, vol. 19, pp. 415–428, 2004.

[17] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[18] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems 14*. Cambridge, MA: MIT Press, 2002.

[19] D. L. Donoho and C. Grimes, "Hessian eigenmaps: New tools for nonlinear dimensionality reduction," *Proc. Nat. Acad. Sci.*, pp. 5591–5596, 2003.

[20] J. Theiler, "Estimating fractal dimension," *J. Opt. Soc. Amer. A*, vol. 7, no. 6, pp. 1055–1073, 1990.

[21] H. S. Greenside, A. Wolf, J. Swift, and T. Pignataro, "Impracticality of a box-counting algorithm for calculating dimensionality of strange attractors," *Phys. Rev. A*, vol. 25, no. 6, pp. 3453–3456, 1983.

[22] J. E. Jackson, *A User's Guide to Principal Components*. New York: Wiley, 1991.

[23] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Proc. 16th IEEE Int. Conf. Tools with Artificial Intelligence*, 2004, pp. 576–584.

[24] C. J. C. Lamoth, A. Daffertshofer, O. G. Meijer, G. L. Moseley, P. I. J. M. Wuisman, and P. J. Beek, "Effects of experimentally induced pain and fear of pain on trunk coordination and back muscle activity during walking," *Clin. Biomech.*, vol. 19, pp. 551–563, 2004.

[25] M. Balasubramanian, E. L. Schwartz, J. B. Tenenbaum, V. de Silva, and J. C. Langford, "The isomap algorithm and topological stability," *Science*, vol. 295, p. 7a, 2002.

[26] J. D. Farmer, E. Ott, and J. A. Yorke, "The dimension of chaotic attractors," *Phys. D*, vol. 7, pp. 153–180, 1983.

[27] J. Guckenheimer, "Dimension estimates for attractors," *Contemporary Math.*, vol. 28, pp. 357–367, 1984.

[28] L.-S. Young, "Dimension, entropy and lyapunov exponents," *Ergod. Th. Dyn. Syst.*, vol. 2, pp. 109–124, 1982.

[29] L. Barreira, Y. Pesin, and J. Schmeling, "Dimension and product structure of hyperbolic measures," *Ann. Math.*, vol. 149, no. 2, pp. 755–783, 1999.

[30] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities: The characterization of strange sets," *Phys. Rev. A*, vol. 33, pp. 1141–1151, 1986.

[31] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed. Hoboken, NJ: Wiley, 2003.

[32] R. H. Clewley, J. M. Guckenheimer, and F. J. Valero-Cuevas, "Estimating degrees of freedom in motor systems," in arXiv:q-bio.QM/0610058 Oct. 2007 [Online]. Available: http://arxiv.org/

[33] B. Scholkopf, A. Smola, and K.-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, pp. 1299–1319, 1998.

Robert H. Clewley received the M.A. degree in computer science from Cambridge University, Cambridge, U.K., in 1995, the M.Sc. degree in nonlinear mathematics from the University of Bath, U.K., in 1997, and the Ph.D. degree in applied mathematics from the University of Bristol, U.K., in 2001.

He is an Assistant Professor in the Department of Mathematics and Statistics, Georgia State University, Atlanta. His research interests focus on methods for characterizing and analyzing effective low dimensionality in dynamical systems models of biological systems.

John M. Guckenheimer received the B.A. degree from Harvard College, Cambridge, MA, in 1966, and the Ph.D. degree in mathematics from the University of California, Berkeley, in 1970.

He is a Professor of mathematics and theoretical and applied mechanics at Cornell University, Ithaca, NY.

Prof. Guckenheimer served as the President of SIAM (1997–1998) and Chair of its activity groups in Life Sciences (2000–2003) and Dynamical Systems (2003–2005). He is also an AAAS Fellow.

Francisco J. Valero-Cuevas (M'99) received the B.S. degree in engineering from Swarthmore College, Swarthmore, PA, in 1988, the M.S. degree in mechanical engineering from Queen's University, Kingston, ON, Canada, in 1991, and the Ph.D. degree in mechanical engineering from Stanford University, Stanford, CA, in 1997.

He is currently an Associate Professor in the Departments of Biomedical Engineering and Biokinesiology and Physical Therapy, University of Southern California, Los Angeles. His research interests focus on combining engineering, robotics, mathematics, and neuroscience to understand organismal and robotic systems for basic science, engineering, and clinical applications.