■ OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

• Laplacian eigenfunctions and neural networks: ...

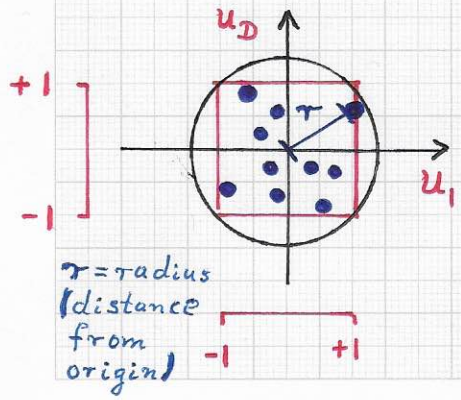
■ CRITERIA FOR SUBDIVISION/SPLITTING. It is necessary to devise a proper, robust and efficiently computable measure for a data set $\{x_i\}$ in D-dimensional space to determine whether the set should be split. This problem relates to multivariate data analysis, i.e., multivariate density functions and distributions. In our case, we must keep in mind specific characteristics of our application: **(i)** We must assume that the number of points in $\{x_i\}$ is "relatively small" when seen in the context of the D-dimensional embedding space. **(ii)** We cannot assume that the finite, discrete point set $\{x_i\}$ satisfies density / distribution characteristics of any specific density function / distribution class. (For example, the Shapiro-Wilk test is a "normality test" that determines whether a finite, discrete sample set is representing / not representing a normal distribution. We cannot employ distribution-specific tests like the Shapiro-Wilk test.) **(iii)** We have to design a measure for the "shape" of the set $\{x_i\}$; specifically, we must determine whether this set is "compact," "dense" or "convex" — using these terms colloquially here. In other words, we must devise a criterion that allows us to decide whether the constructed (minimal) bounding box of $\{x_i\}$ suffices or one should split the set.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

• Laplacian eigenfunctions and neural networks:...

Statistical moments can be used in the univariate/1-dimensional setting to characterize the "shape" of a distribution. We briefly review the necessary definitions of low-order moments for the univariate case. Since known or to-be-constructed generalizations of moments in the multivariate case could be considered or developed, one might consider using such generalized moments. A much simpler, more efficient and straightforward approach would use the well-known univariate moments "directly" in the multivariate setting. This approach would have to use one parameter/variable on which a density/distibution in a D-dimensional space depends. The distance (radius) from the center of the sample points can serve as this one parameter/variable. In the context of described linear coordinate transformation applied to a point set, the use of univariate moments can be understood as follows:



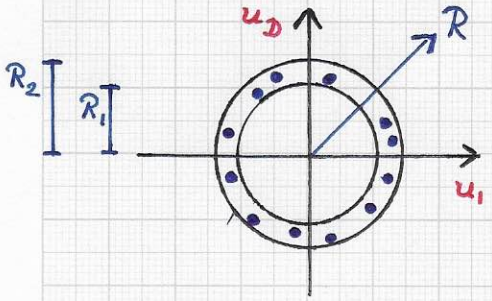
The left figure illustrates the fact that we can use the u1-tuples of the points – all lying in the normalized (bounding) hyper-box $[-1,1]^D$ with corresponding circumscribed hyper-sphere (-ball) $\sum_{d=1}^{D} u_d^2 = 1 \ (\leq 1)$.

Density/distribution functions can be defined relative to r.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

• Laplacian eigenfunctions and neural networks: ... In order to compute density/distribution function estimates, we must utilize the general formula for the (hyper-) volume of a ball with radius **R** in **D** dimensions. First, we must define a simple method for estimating point density as a function depending on radius R. The left figure sketches the underlying concept of a density estimation scheme: For two (hyper-) balls in D-dimensional space, with both balls having the origin as center and radii R_1 and R_2, one can count the number of points lying inside the ball with the larger radius (R_2) and not simultaneously also lying inside the ball with the smaller radius (R_1); further, one can compute the (hyper-) volume bounded by the two (hyper-) spheres with radii R_1 and R_2 — and density can be estimated as the number of points counted divided by the computed (hyper-) volume. The figure illustrates the case $D=2$. Here, 12 points '•' lie inside the 2-ball with radius R_2 but not inside the 2-ball with radius R_1 ($R_1 < R_2$). Thus, the resulting density estimate is $12/(\pi R_2^2 - \pi R_1^2) = 12/(\pi(R_2^2 - R_1^2))$. This value can be associated with the interval $[R_1, R_2)$ or radius $(R_1+R_2)/2$ as density estimate.



•••

■ OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

• **Laplacian eigenfunctions and neural networks:** ... Of course, in the D-dimensional case, we must use the general formula for the (hyper-)volume of D-balls. It is desirable to consider two formulae for the (hyper-) volume of a D-ball, one formula for even values of D and one formula for odd values of D; they are:

$$V = \frac{1}{(D/2)!}\pi^{D/2} R^D \quad , D \in \{0, 2, 4, \ldots\} ;$$

$$V = \frac{2^{\lceil D/2 \rceil}}{D!!}\pi^{\lfloor D/2 \rfloor} R^D \quad , D \in \{1, 3, 5, \ldots\} .$$

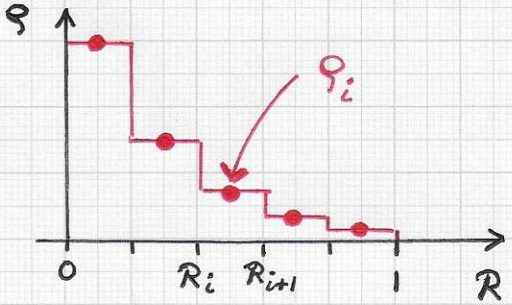
Here, $D \in \{0, 1, 2, \ldots\}$, $n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$ (where $0! = 1$), $n!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (n-2) \cdot n$ (where n is odd) and the notation "$\lfloor\ \rfloor$" and "$\lceil\ \rceil$" is used for floor and ceiling. For example, the volume of a 10-ball is given as $V(10) = 1/5!\ \pi^5 R^{10} = 1/120\ \pi^5 R^{10}$ and the volume of an 11-ball is given as $V(11) = 2^6/11!!\ \pi^5 R^{11} = 64/10395\ \pi^5 R^{11}$. For $D=0$, $D=1$, $D=2$ and $D=3$, one obtains the well-known formulae $V(0) = 1$, $V(1) = 2R$, $V(2) = \pi R^2$ and $V(3) = \frac{4}{3}\pi R^3$.

• **Note.** In our setting and application, all points lie in the unit (= radius-one) D-dimensional ball, i.e., $R \leq 1$ holds for all points (after application of the described linear coordinate transformation). Values of D can be "relatively large," much larger than three. The number of points in $\{u_i\}$ is likely in the range 10^2 to 10^6.

• • •

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks: ... Further, one must keep in mind that a "large number of points in a very-high-dimensional D-ball" does not necessarily lead to high point density. Thus, point density estimates "are better" for lower values of D. The volume $V = V(D)$ of a D-ball can be written as $V(D) = c_D \cdot R^D$, where the polynomial coefficient c_D depends on the dimension D, as defined on the previous page. In order to calculate a discontinuous, piecewise constant approximation of point density $\varrho = \varrho(R)$, we subdivide the unit interval $[0, 1]$ into sub-intervals $[R_i, R_{i+1})$, where $R_0 = 0$, $R_i < R_{i+1}$, $R_N = 1$ and $\cup_i [R_i, R_{i+1}) = [0, 1)$. The figure shows that we associate the density value ϱ_i with sub-interval $[R_i, R_{i+1})$. Assuming we use $(N-1)$ sub-intervals, we must comute interval-specific density estimates for $[R_0 = 0, R_1), \ldots, [R_{N-1}, R_N = 1)$, i.e., $\varrho_0, \ldots, \varrho_{N-1}$. Specifically, we must determine the number of points that satisfy the condition $R_i^2 \le \sum_{d=1}^{D} u_d^2 < R_{i+1}^2$ (to lie in the region), $N_{[R_i, R_{i+1})}$, associated with sub-interval $[R_i, R_{i+1})$; further, we must compute the (hyper-) volume of the respective region, i.e., $V_{[R_i, R_{i+1})} = c_D \cdot (R_{i+1}^D - R_i^D)$. Thus, the resulting density is



$$\varrho_i = N_{[R_i, R_{i+1})} / V_{[R_i, R_{i+1})} = N_i / V_i .$$

...