·Poor understanding of what can be done to process high dimensional data .

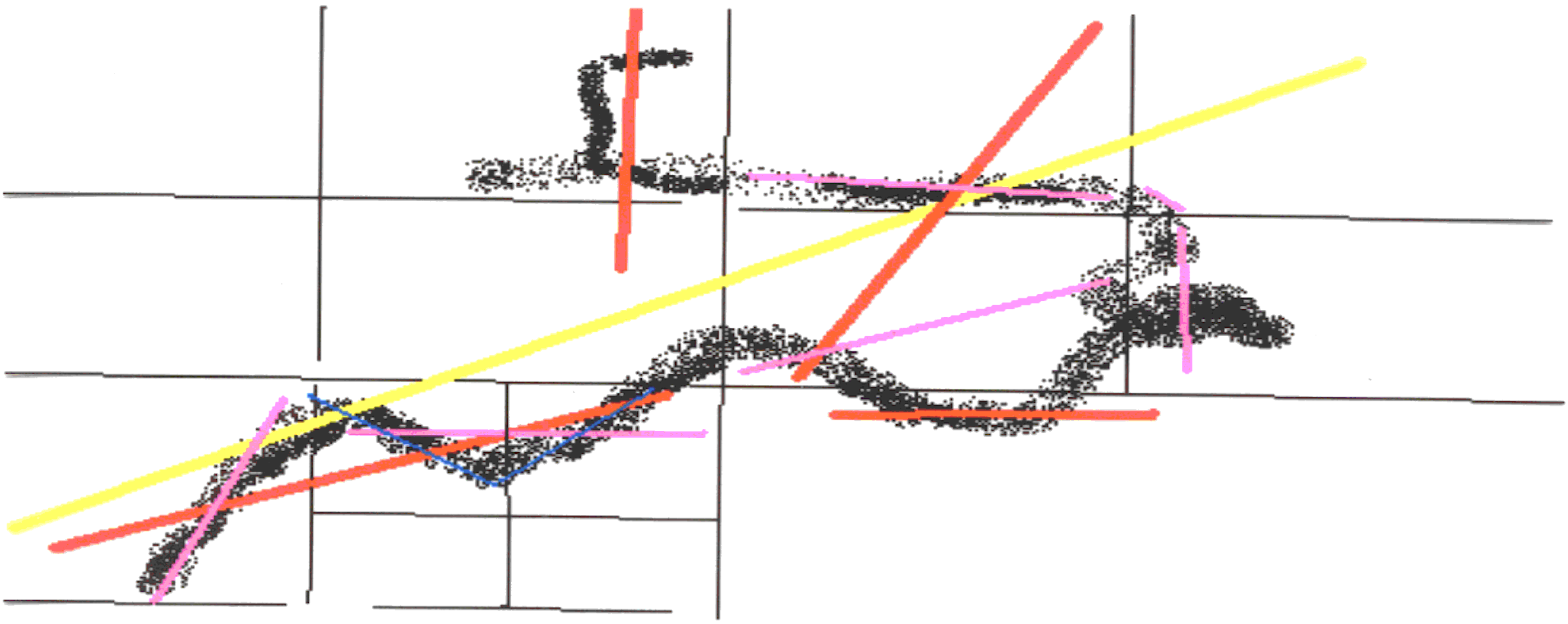
· Empirical Data (sets in high d) used to do prediction or estimation are not rich enough,and need to be modeled .

·We don't have quantitative methods to build dimensional nonparametric models ,algorithms like neural nets or self organizing maps might work in practice but are nor supported by estimation.

·No effective algorithms to approximate function in high dimensions ,all are exponential in the dimension.

· The main tool for dimensional reduction in linear problems i the ***Singular Value Decomposition***,it allows effective rank computations , but is not coupled generally to FFT type fast algorithms. Nothing like this exists for nonlinear maps.

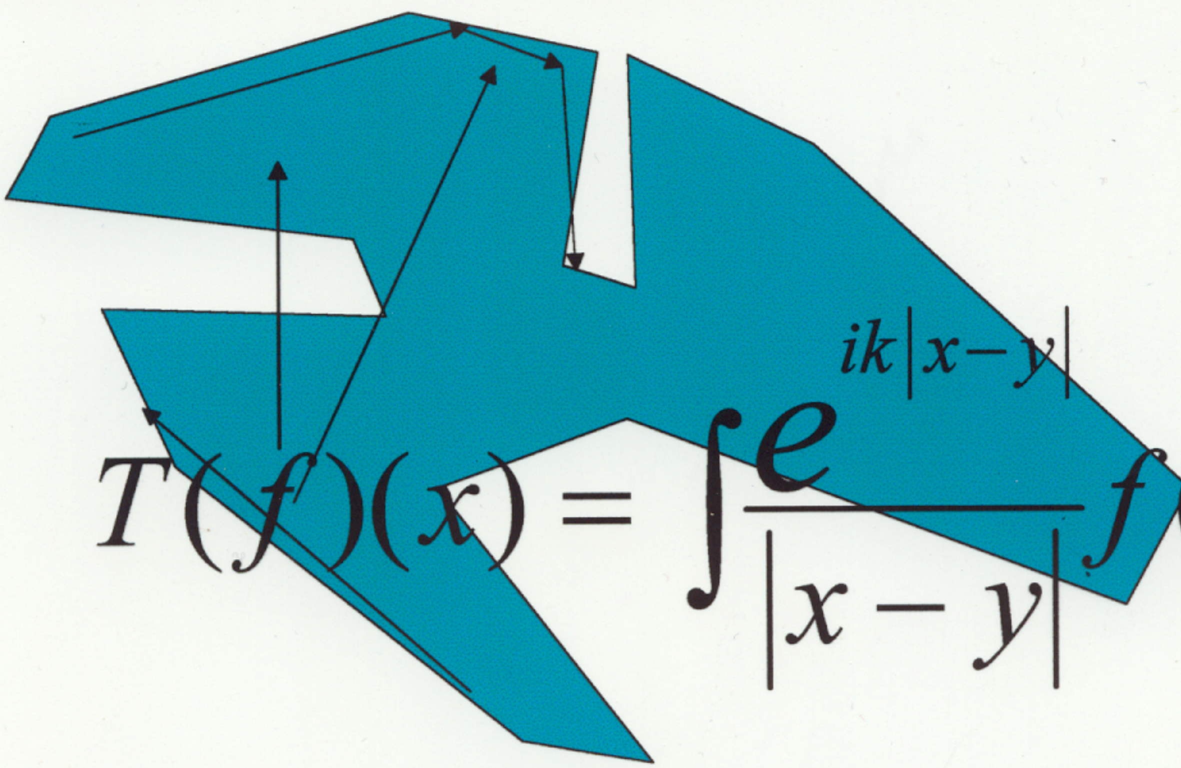
·Localization of SVD can lead to fast effective algorithms



A multiscale approximation of a set in the plane (P.Jones)
The max of ratios between the side of a square and the sum of deviations on all subsquares controls the bi Lipshitz constant of a one dimensional parametrization of most of the set

A data driven binary transcription of a set

Acoustic scattering off object requires detailed effective
Field interactions between regions on boundary.

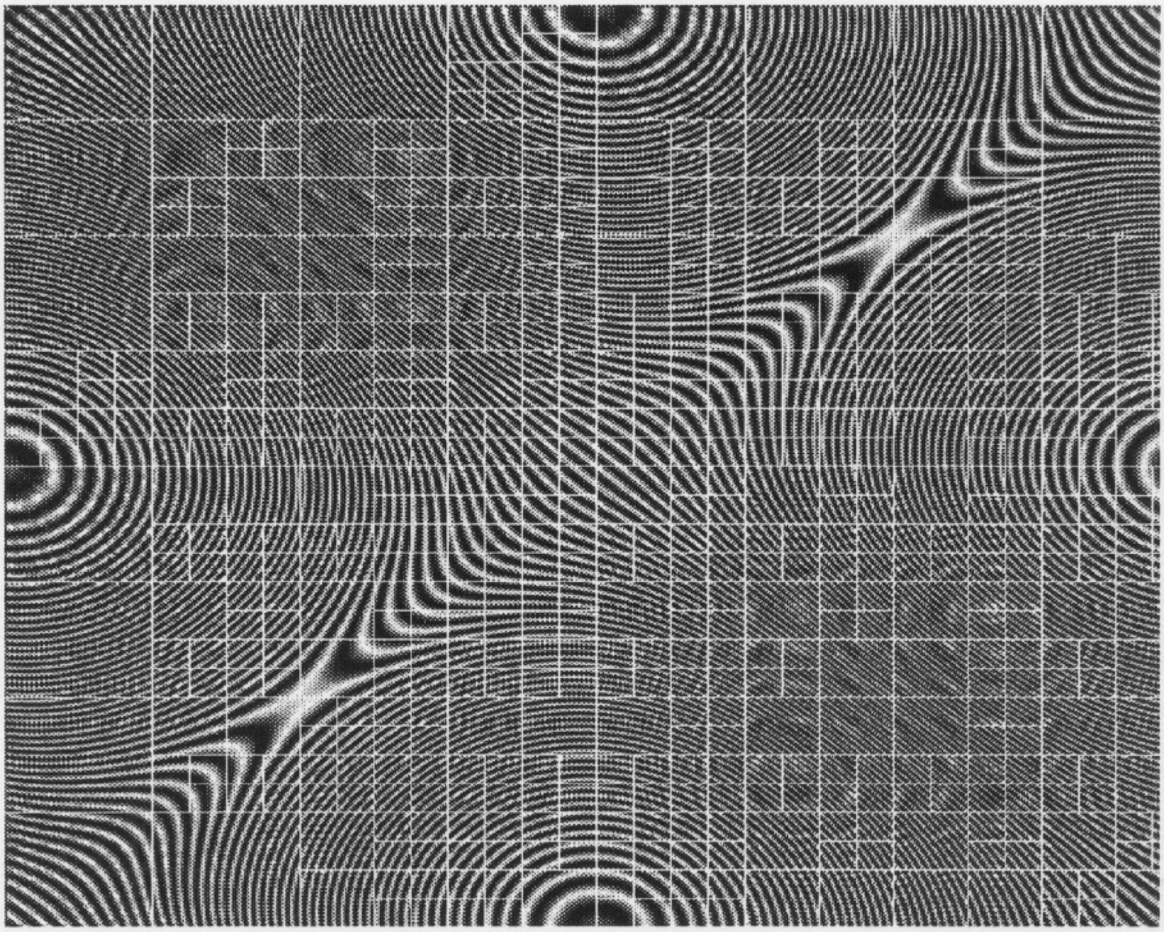


$$T(f)(x) = \int \frac{e^{ik|x-y|}}{|x-y|} f(y)$$

The first approximation is given by geometric optics ,or Billiards and i
Obtained automatically through orchestration.

Acoustic scattering matrix off an ellipse ,while dense ,the number of parameters (features) needed to describe it is small.each box encapsulates geometric optics interaction.

The same analysis could be otained by local SVD analysis to track rank of interactions The rank is one at the geometric optic level



BASES OF FUNCTIONS

1. EXPRESS ALL FUNCTIONS AS SUMS OF (COEFFICIENTS TIMES) BASIS FUNCTIONS
2. ANALYSIS: THROW OUT SOME COEFFICIENTS (COMPRESSION)
3. SYNTHESIS: GENERATE FUNCTIONS (SIGNALS) BY GENERATING COEFFICIENTS

DATA SETS ARE DIFFERENT:

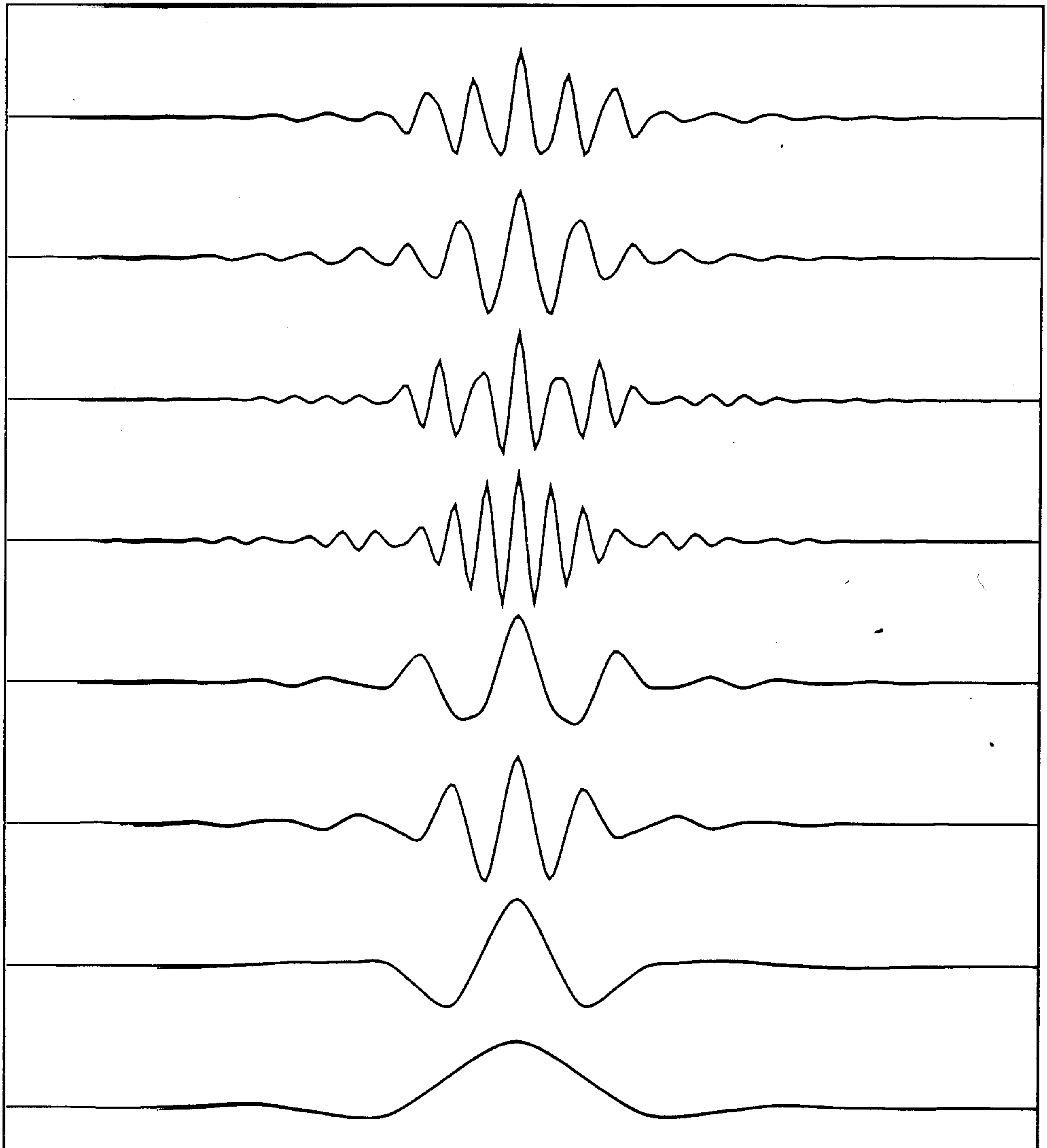
NOT FUNCTIONS (PROBABILITY DISTRIBUTIONS)

LIVE IN HIGHER DIMENSIONAL SPACES

BASIS FUNCTIONS DON'T HAVE IMMEDIATE ANALOGUE

SIMILARITIES:

STUDY BY PROJECTING ONTO LOW DIMENSIONAL SPACES



DICTIONARY

wavelets	β numbers
$a_{j,k}$ for function f	$\beta_{Q_{j,k}}$ for set K
analysis and synthesis of the function f	analysis and synthesis of curve $\Gamma \supseteq K$
$\|f\|^2 = \sum \lvert a_{j,k} \rvert^2$	$l(\Gamma) \sim \sum \beta_Q^2 \cdot l(Q)$
square function $W_\psi(x)^2$	$J(x)$

II.2 Anisotropy Scaling Analysis

Peter Jones, *Rectifiable Sets and Travelling Salesman Problem*
Guy David and Stephen Semmes, *Analysis of and on Uniformly Rectifiable Sets*

Question

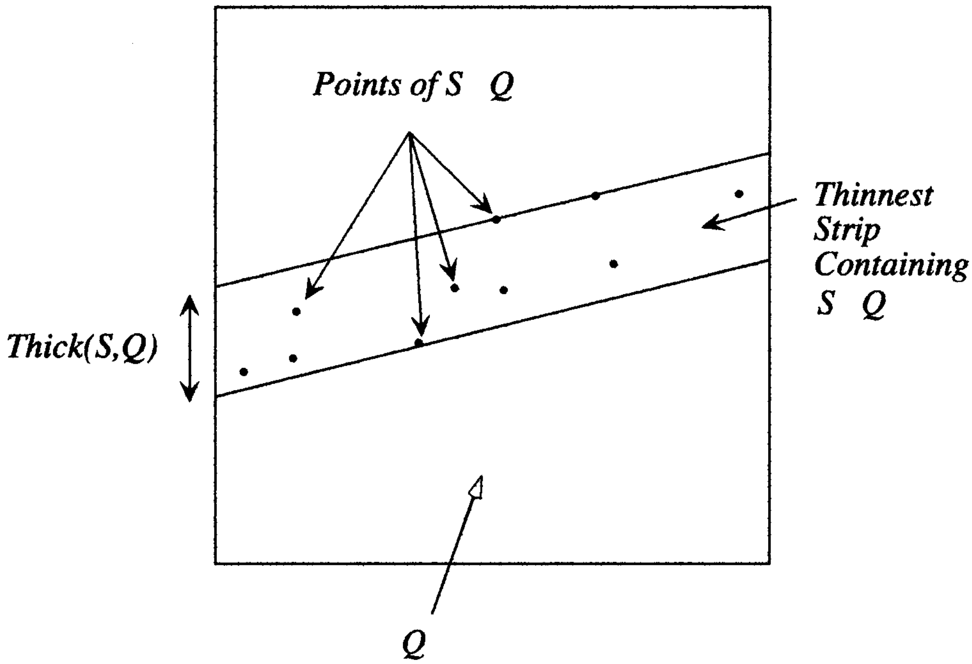
$S \subset [0,1]^2$ – is S a subset of rectifiable curve?

Analysis Tool: Multiscale Thickness

$$Thick(S,Q) = \text{Width of Thinnest Strip} S \cap 3Q$$

where Q is a dyadic square.

$$\beta(S,Q) = Thick(S,Q)/\ell(Q)$$



Jones Travelling Salesman Theorem. *There exists a finite-length curve* Γ *containing* S *if and only if*

$$\sum_Q \beta^2(S,Q)/\ell(Q) < \infty$$

I.E. the set is anisotropically "thin" at most scales and locations.

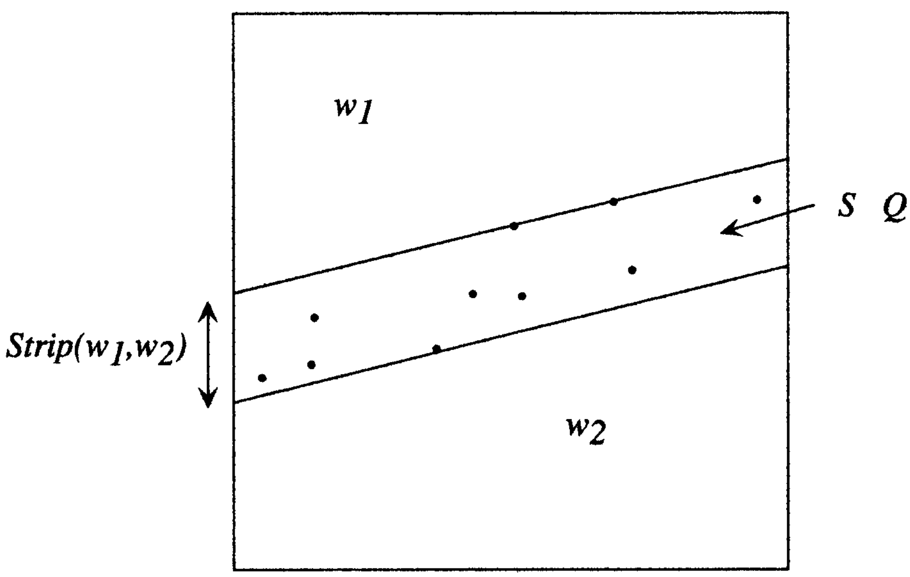
Empirical Thickness Analysis

$I(i_1, i_2)$ black-or-white pixel-level image

$$\hat{\beta}(I, Q) = \widehat{\mathrm{Thick}}(\{I = 1\}, Q)/\ell(Q)$$

Relation to Wedgelets:

If two disjoint wedgelets w_1, w_2 associated to a square Q each contain zero mass, they define a complementary strip $Strip(w_1, w_2)$



Definition. $\widehat{\mathrm{Thick}}(\{I = 1\}, Q)$ is thinnest of all such strips.

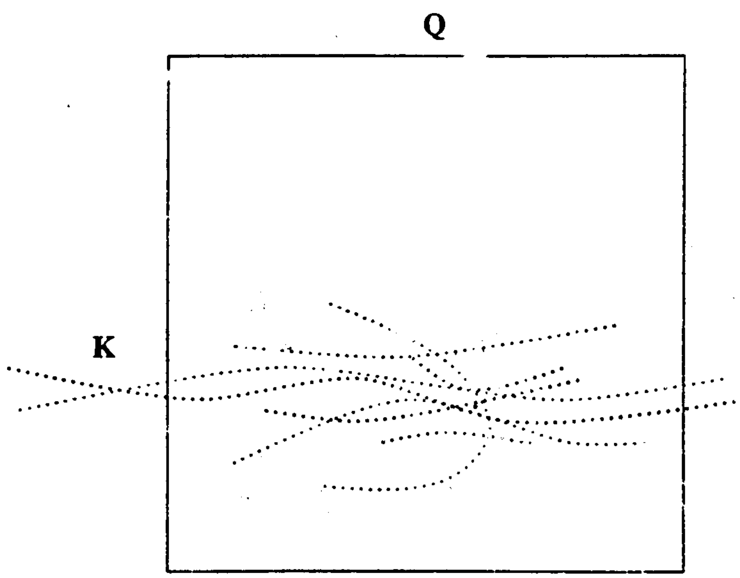
ONE-DIMENSIONAL GEOMETRIC TRANSCRIPTIONS

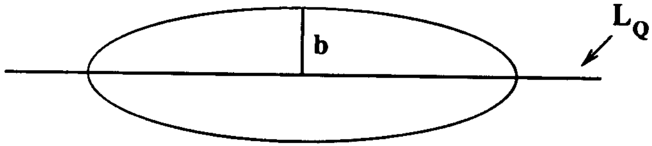
Consider a data set $K \subseteq \mathbb{R}^n$

β number for a cube Q

$L_Q =$ best l_2 approximating line for $K \bigcap Q$

$$\beta_Q = \frac{l_2\text{-average distance of } K \bigcap Q \text{ from } L_Q}{l(Q)}$$





$$\beta_Q = \frac{b}{l(Q)}$$

$\varepsilon \leq 3 \leq$ "1"

d-DIMENSIONAL GEOMETRIC TRANSCRIPTIONS

β number for a cube Q

D_Q = best l_2 approximating d-plane for $K \bigcap Q$

$$\beta_Q = \frac{l_2\text{-average distance of } K \bigcap Q \text{ from } D_Q}{l(Q)}$$

- β_Q is determined by the singular values of the data matrix.

THE COMPUTATION OF β_Q

- Let $N = |K \cap Q|$.

- Let A be an $N \times n$ matrix whose rows are the points of $K \cap Q$.

- The axes of the ellipsoid are the singular vectors of A.

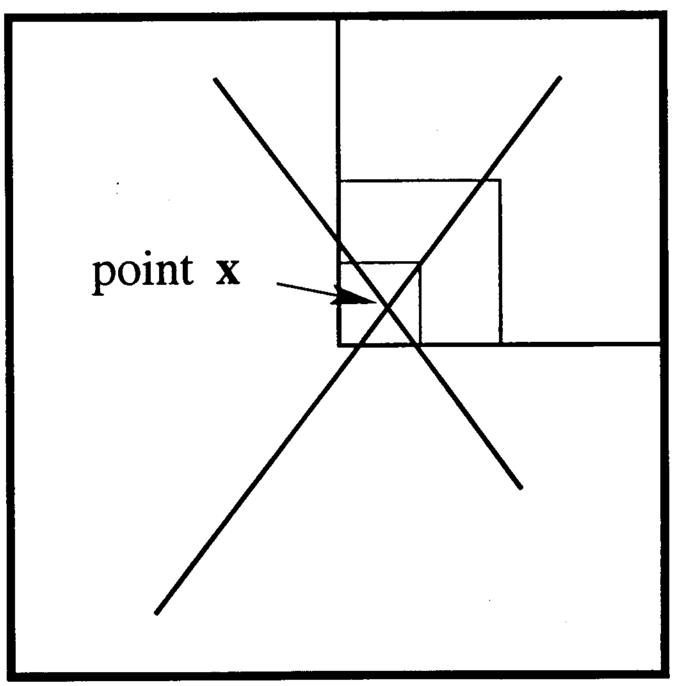
- β_Q is determined by the singular values of A.

- β_Q is computed in $O(n^2 \cdot N + n^3)$ operations.

JONES' FUNCTION FOR A DYADIC GRID

$$J(x) = \sum_j \beta^2_{Q_j},$$

Q_j : dyadic cube containing x, $l(Q_j) = 2^{-j}$.



$$J(x) = \sum_j \beta_{Q_j}^2$$

Q_j : dyadic cube containing x, $l(Q_j) = 2^{-j}$

J(x) measures rectifiability of K around x:

If $J(x) \leq M$ for all $x \in K$, then a portion of K is contained in a curve, whose length is controlled by M.

PROBLEM:

Approximate $K \subseteq \mathbb{R}^n$ by curves

APPROACH:

Approximation by lines at different scales

β number of a cube Q:

$L_Q =$ best l_2 approximating line for $K \bigcap 3Q$

$$\beta_Q = \frac{l_2\text{-average distance of } K \bigcap 3Q \text{ from } L_Q}{\sqrt{l(Q)}}$$

DESCRIPTION OF THE ALGORITHM

One-stage algorithm for a set K:

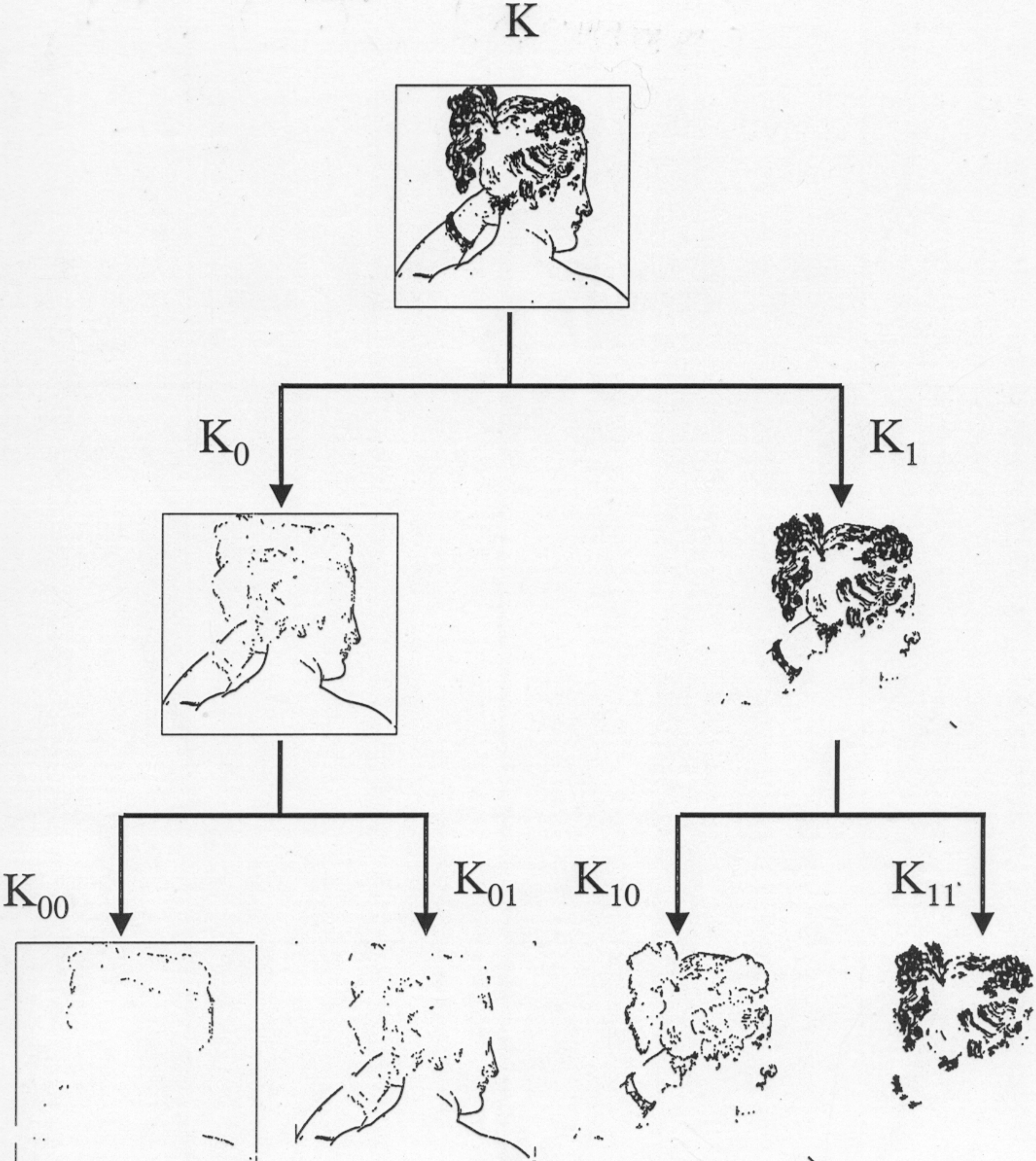
1. Compute $J(x)$ for all $x \in K$.

2. Find
$$M = \max_{x \in K} J(x)$$
$$m = \min_{x \in K} J(x)$$

3. Construct
$$K_0 \equiv \{ x \mid m \leq J(x) < \frac{m+M}{2} \}$$
$$K_1 \equiv \{ x \mid \frac{m+M}{2} \leq J(x) \leq M \}$$

Multistage algorithm:

Apply repeatedly the one-stage algorithm to K, K_0, K_1, ...


K
K_0
K_1
K_{00}
K_{01}
K_{10}
K_{11}

β_∞ number for a cube Q:

$$\beta_{\infty Q} = \frac{\text{smallest width of a strip containing } K \cap Q}{l(Q)}$$

Theorem (P. W. Jones, 1990)

Let K be a subset of $\mathbb{R}^2$. K is contained in a curve with finite length $\Longleftrightarrow$ diam(K) and $\sum \beta_Q^2 \cdot l(Q)$ are finite.
Moreover, the length of the shortest curve containing K is comparable to
diam$(K) + \sum \beta_Q^2 \cdot l(Q)$. $(\beta = \beta_\infty)$

Theorem (C. Bishop, P. W. Jones, 1990)

If K is a bounded set in $\mathbb{R}^n$ and if $J(x) \leq M$ for all $x \in K$, then K is contained in a curve of length not exceeding $c_1 e^{c_2 M}$diam(K).

Th. (J, Lerman)

$\exists\ c_0, c_1$ (Universal)

s.t. if we set

$$I(\mu) = \int \exp\left\{c_1 \sum_{Q \ni x} \beta^2(Q)\right\} d\mu(x)$$

then $\exists$ "good" d-dim "manifold" S s.t.

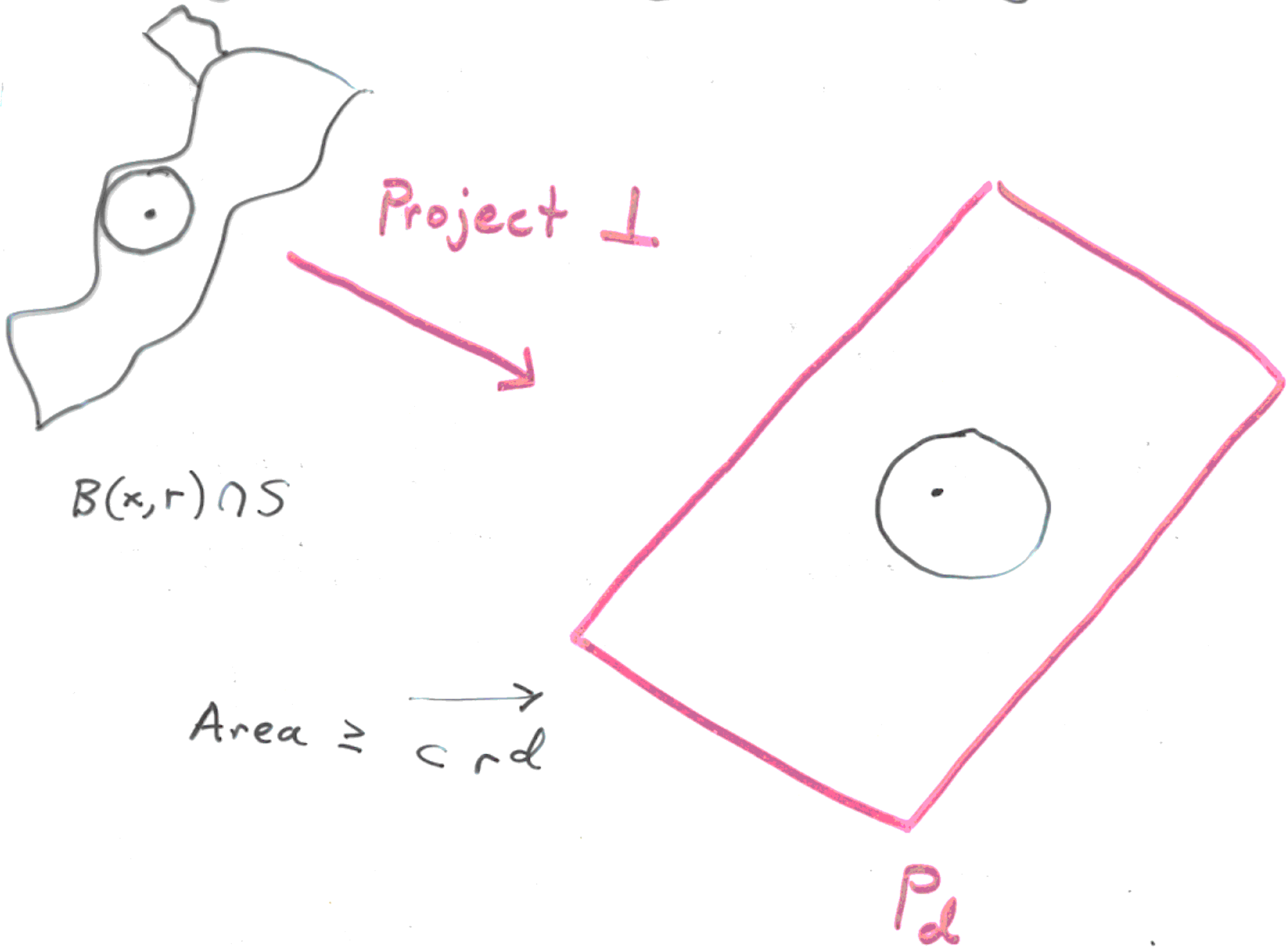
1) $\text{Area}_d(S) \le c_0\, I(\mu)^{+1}$

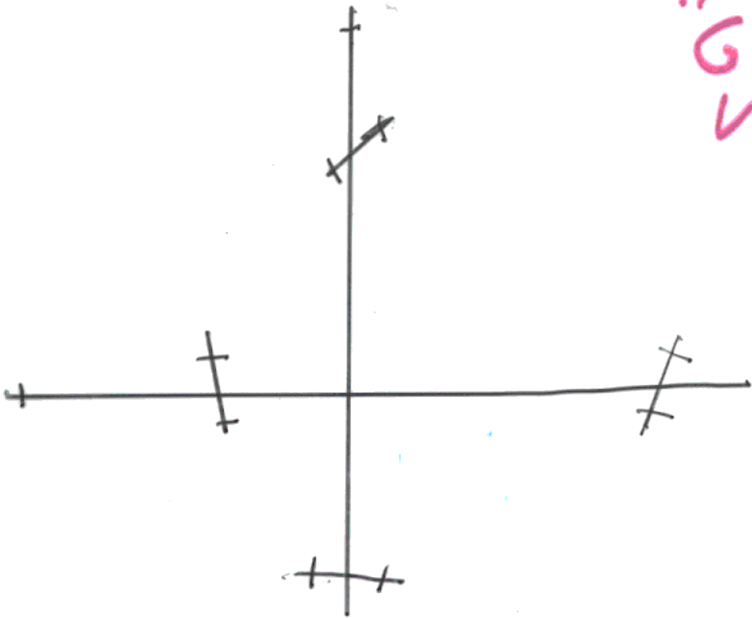
2) $\mu(S) \ge \frac{1}{c_0} I(\mu)^{-1}$

☆ Everything sharp

If we replace c_1 by $c_2 \ll 1$ and $I_{c_2}(\mu) \le C_3$, No Conclusion Possible!

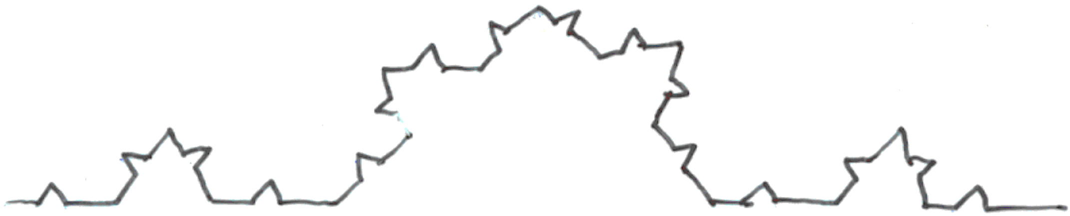
These Have BIG PROJECTIONS





"Glued" Version

Length Must Grow



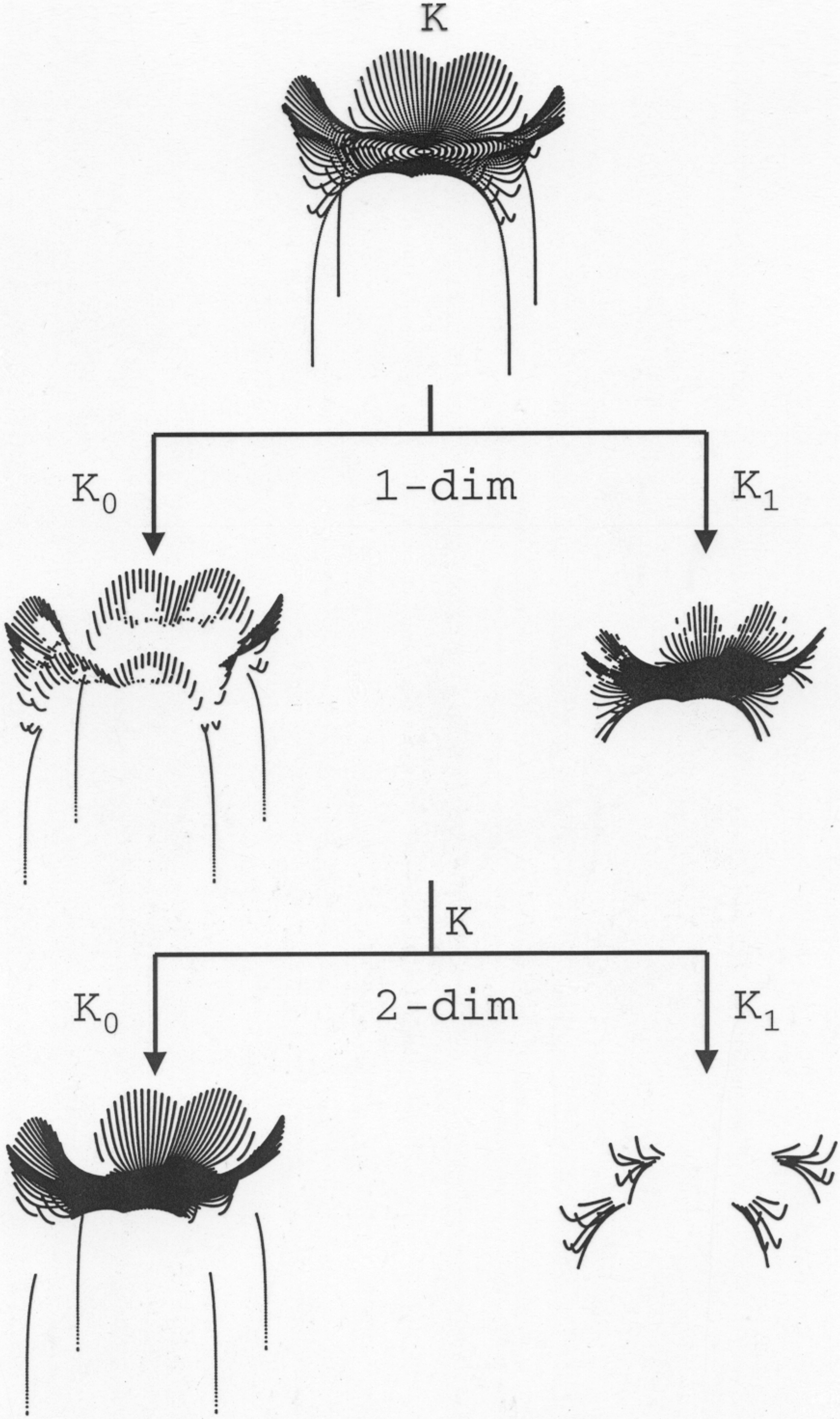
Run N stages in Von Koch curve

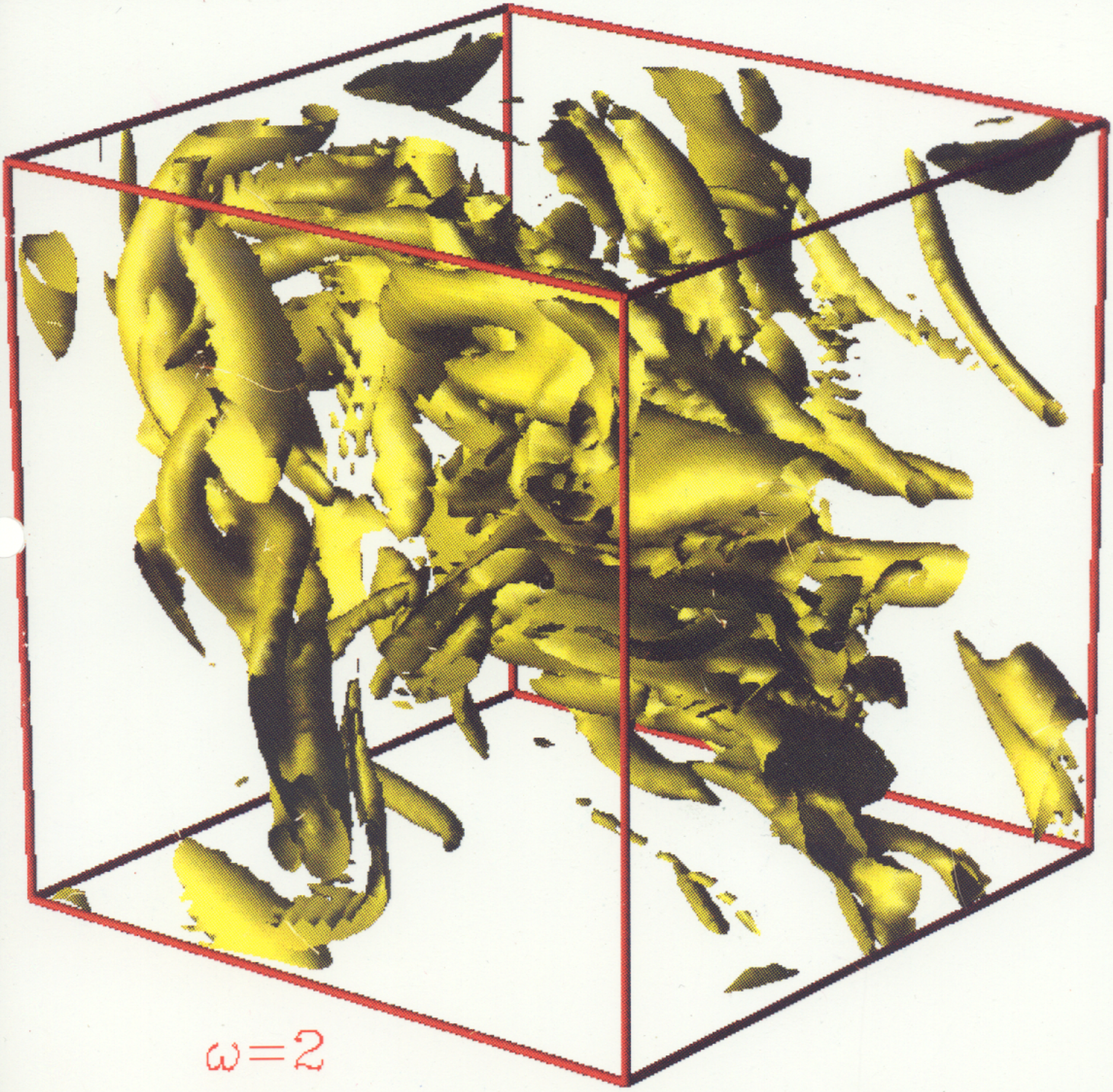
$$\sum_{Q \ni x} \beta^2(Q) \sim N$$

$$\text{length} = \left(4/3\right)^N = e^{(\log 4/3)\cdot N}$$

$$\sim \int e^{c \sum \beta^2(Q)} \, d\mu$$

(As in Conclusion)


K
K0
1-dim
K1
K
K0
2-dim
K1


ω=2