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.

Model Problems Related to β

Jitendra Malik
David Mumford

1. Cauchy Integrals on Lipschitz Curves

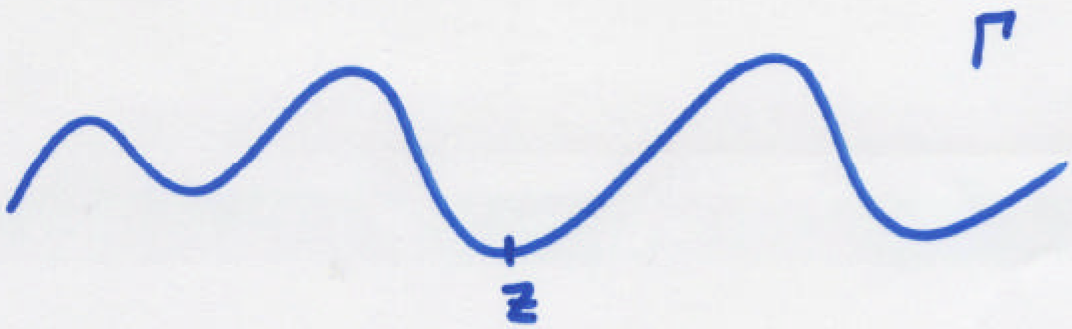
2. Analytic Capacity

3. Estimates for Brownian Motion

4. Estimates (Non-Sharp!) for Dimension

5. Mumford-Shah Conjecture (Work of David-Semmes on K, $\mathcal{H}^1(K)$, ...)

Model Problems: Cauchy Integ. on Lip Curves, Analyt. Cap. (Tolsa, NTV), Sharp estimates for ω (Schw, Brownian Motion)



$$\mathcal{C}f(z) = \frac{1}{2\pi i}\int_{\Gamma} \frac{f(w)\,dw}{w-z}$$

Theorem (Calderón, C-M-M)

If Γ is a Lipschitz curve, then

$$\|\mathcal{C}f\|_{L^2(\Gamma)} \leq A\|f\|_{L^2(\Gamma)}$$

A depends on Lipschitz Const.

$\mathcal{C} \approx H$ (diagonal) + Off Diagonal

Off Diag. Controlled by $\sum \beta^2(Q)\,\ell(Q)$

Def. $\gamma(K) > 0 \iff$

$\exists\, F \not\equiv$ Constant,

F bounded, analytic on K^c

Surprise: This is an L^2 problem!

Critical Case: $\dim(K) = 1$

$\ell(K) = 0 \Rightarrow \gamma(K) = 0$

1. (David) If $0 < \ell(K) < +\infty$,

$\gamma(K) = 0 \iff \ell(K \cap \Gamma) = 0$

for all curves

Γ, $\ell(\Gamma) < +\infty$

Melnikov's Philosophy

M. Conjecture: $\gamma(K) > 0 \iff$
$\exists\, \mu \in Pr(K)$ s.t.

(1) $\mu(D(x,r)) \le Ar$

(2) $\iiint c^2(x,y,z)\, d\mu(x)\, d\mu(y)\, d\mu(z) < \infty$

Note: C.Z. theory shows $\Leftarrow$
Hard Part is to show $\Rightarrow$

Thm 1. says MC is true when
K is a continuum.
(Then $\gamma(K) > 0$ because
of Riemann mapping

$$F : S^2 \setminus K \longrightarrow \mathbb{D}$$

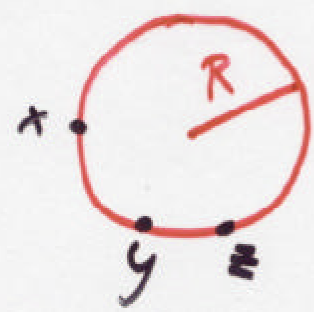
$\uparrow$ unbounded component
unit disk

Tolsa's Theorem: $\gamma(K) > 0$

$\Longleftrightarrow \exists\ \mu \in Pr(K)$:

$$\mu(D(x,r)) \le A r$$

$$\int_K\int_K\int_K c^2(x,y,z)\, d\mu(x)\, d\mu(y)\, d\mu(z) < \infty$$



$c = 1/R$

Problem: Given $\gamma(K) > 0$ (or μ as above),

<u>CONSTRUCT</u> $f \in H^\infty(K^c)$

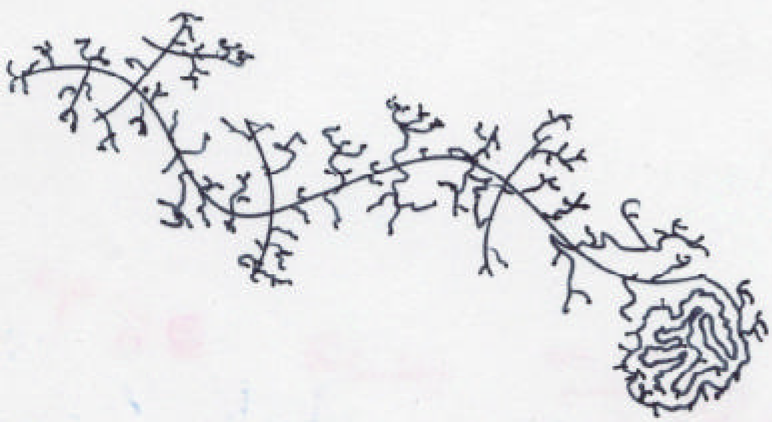
Theorem 2. (J)

$K \subset \mathbb{C}$ a continuum,

Diameter$(K) = 1$

$\Rightarrow \exists\, \mu \in Pr(K)$,

$$\left\| \mu * \frac{1}{z} \right\|_{H^\infty(\mathbb{C} \setminus K)} \leq C_0$$



$H^\infty \leftrightarrow$ Holomorphic, Bounded

Sup Norm

Nazarov - Treil - Volberg '98

Found the correct extension of the $\{\psi_I\}$ method (C-J-S) for non-doubling measures.

A. Volberg will give a lecture series on this plus more recent advances such as:

$$K \subset \mathbb{R}^n$$

When is there $\Delta u = 0$ on K^c, $\nabla u \in L^\infty(K^c)$?

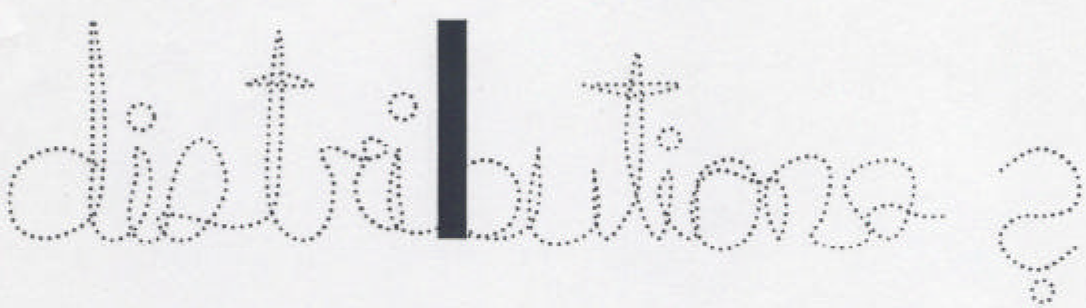
Answer: $\iff \exists \ \mu \in Pr(K)$

$$\mu * \vec{K} \in L^\infty$$

↑ this is a kernel! no relation to the set.

Low-dimensional data sets

- Mixed dimensionalities.
- Non-smooth structure.
- Noise.
- Dependence on the distribution of samples.



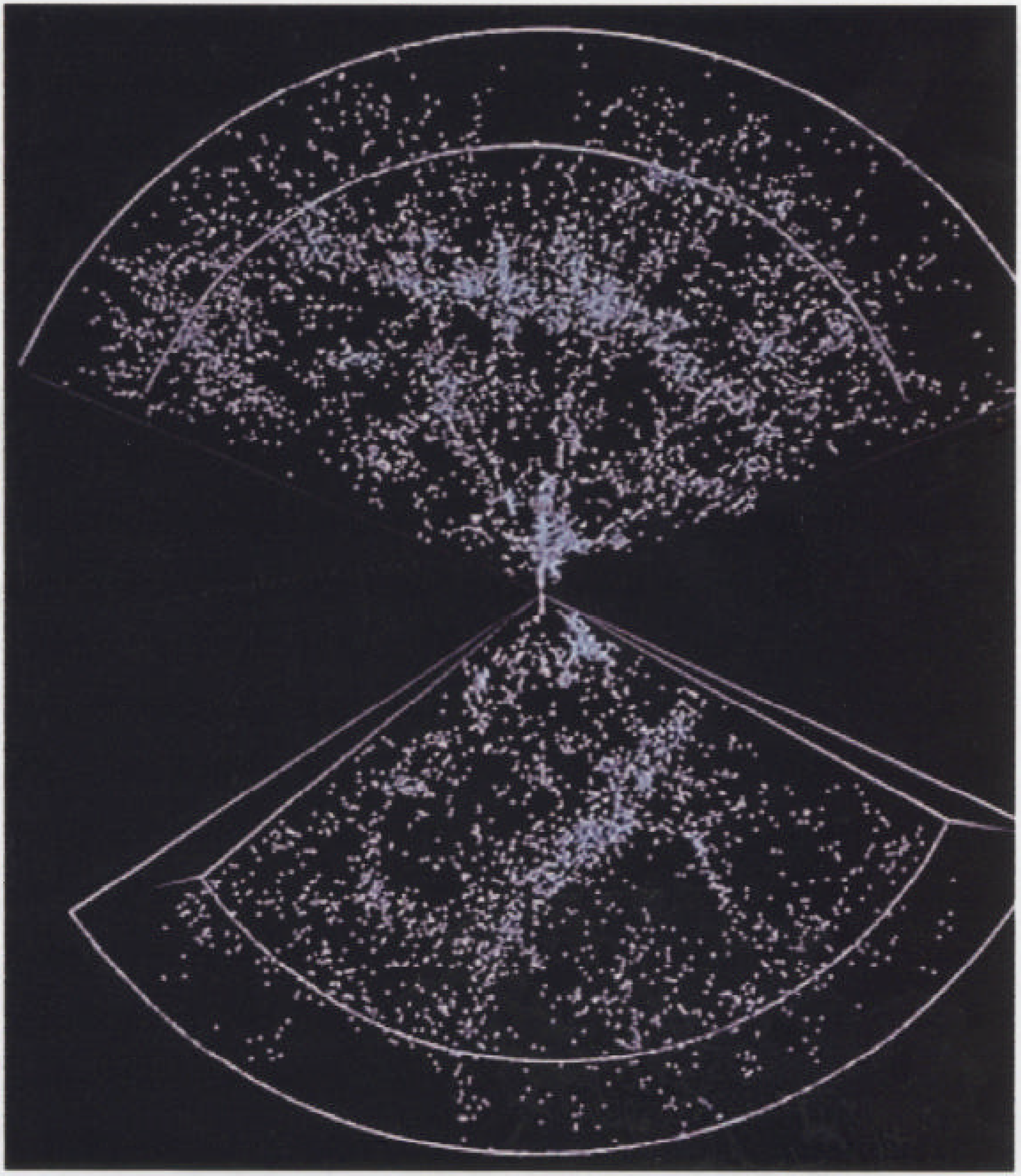
MOTIVATION

Goal

Efficient description of low dimensional data sets in $\mathbb{R}^n$:

- Fast to compute.
- Insensitive to noise.
- Useful in analysis and synthesis of data sets.

arranged in thin sheets or long filaments. The longest sheet detect the "Great Wall," extends hundreds of millions of light years acr maps.



North and South, Sheets and Voids

(Courtesy: Margaret J. Geller and Emilio E. Falco, Harvard-Smithsoni Astrophysics. Credits: Geller, da Costa, Huchra, and Falco.) Large-scale structure in the universe in the northern and south

CURVE-LIKE SETS

Heuristic definition

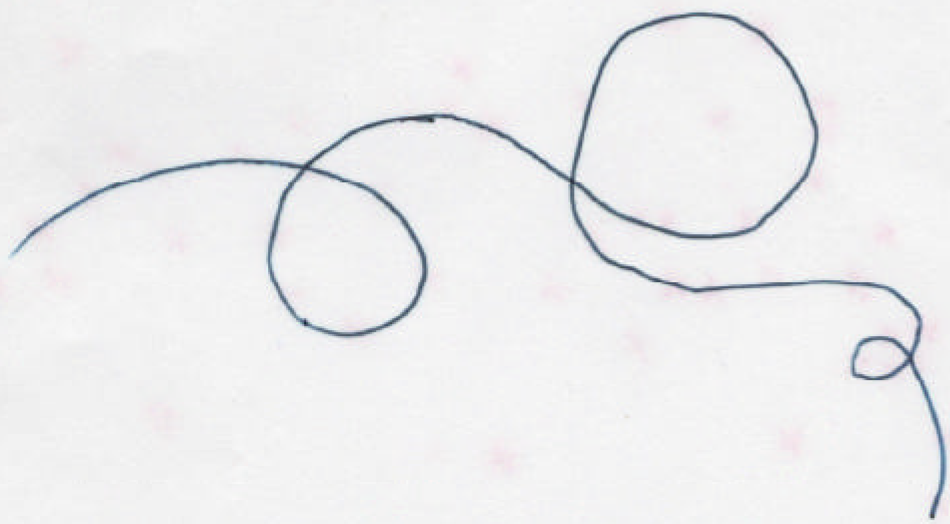
There exists a curve of short length containing a large fraction of the set.

Related question: approximate TSP

Construct a nearly shortest curve containing the given set.

Another definition

The set is well approximated by lines at different scales and locations.



∃ <u>Short</u> Curve Through <u>Many</u> Points?

Approximate TSP

56

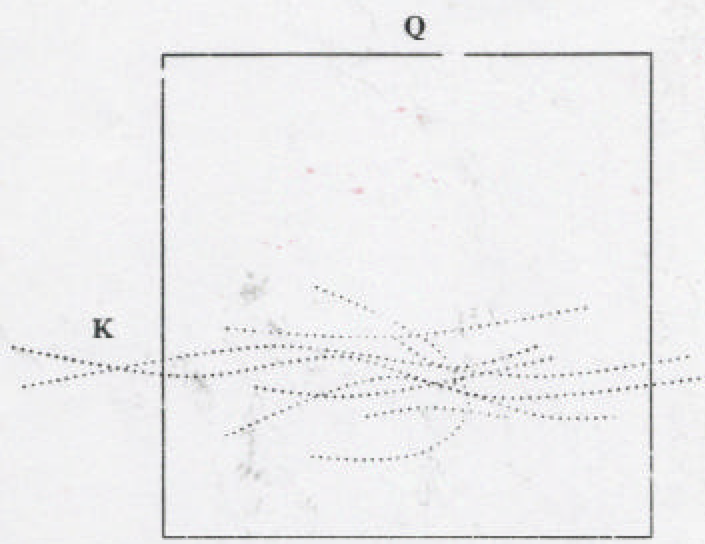
ONE-DIMENSIONAL GEOMETRIC TRANSCRIPTIONS

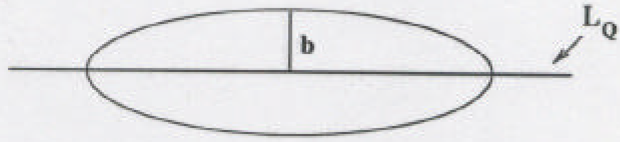
Consider a data set $K \subseteq \mathbb{R}^n$

β number for a cube Q

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

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





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

$0 \leq \beta \leq$ "1"

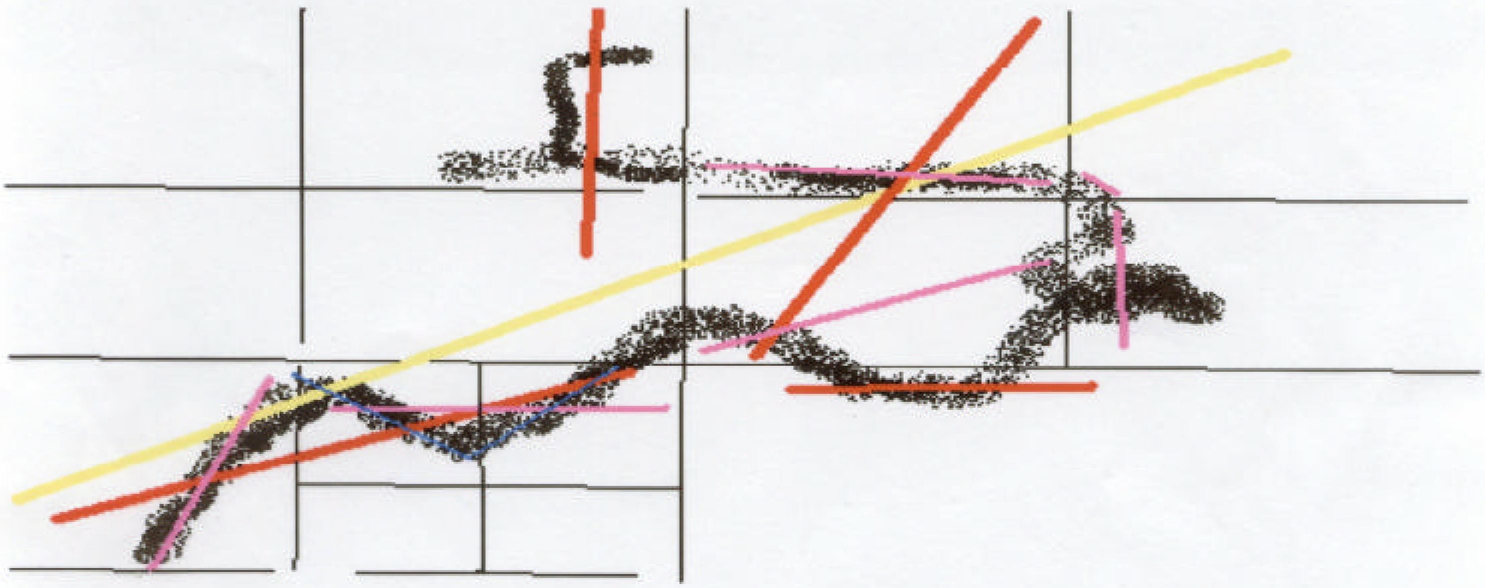
d-DIMENSIONAL GEOMETRIC TRANSCRIPTIONS

β number for a cube Q

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

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

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



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

61

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

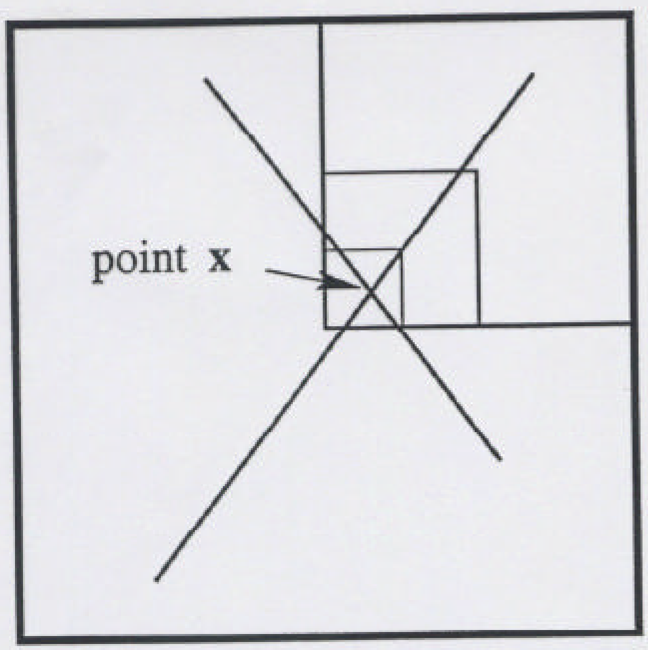
Q_j : dyadic cube containing x, $l(Q_j) = 2^{-j}$

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.

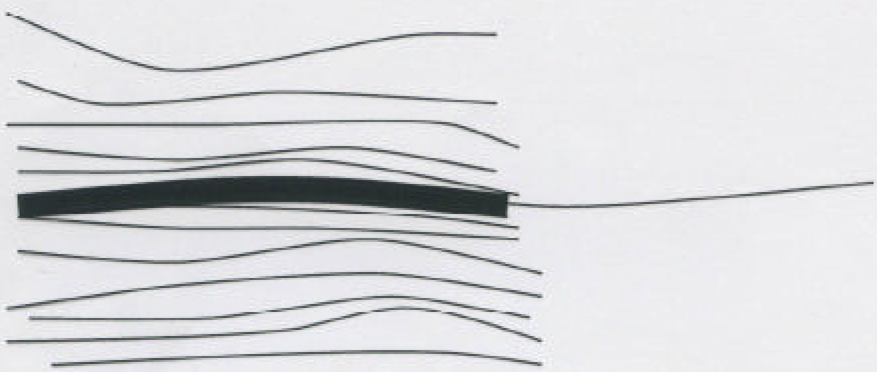
$$J(x) = \sum_j \beta_{Q_j}^2,$$

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

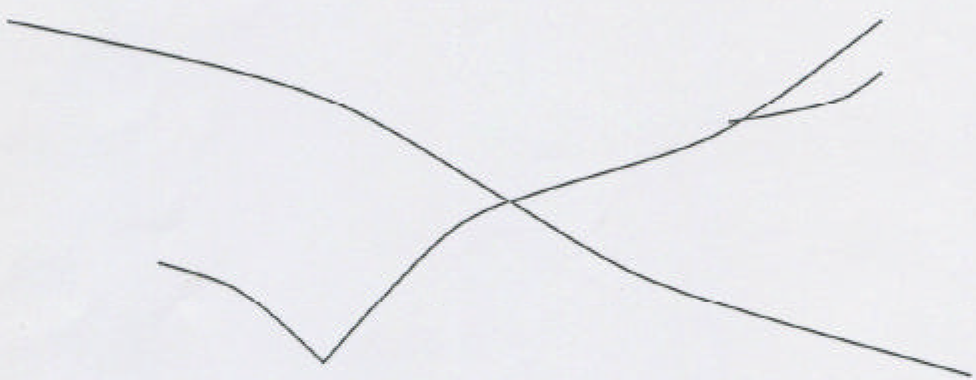


Where is J(x) large ?

- Higher dimensional structure.



- Non-smooth structure (crossings, corners).



THEOREM (Lerman)

Let K be a set in $\mathbb{R}^n$ with $\tilde{J}(x) \leq M$ for all $x \in K$. Then there exists a curve Γ such that

$$l(\Gamma) \leq c_1 e^{c_2 M} \cdot \mathrm{diam}(K)$$

and

$$|\Gamma \cap K| \geq \frac{1}{c_1} e^{-c_2 M} \cdot |K|.$$

Remarks:

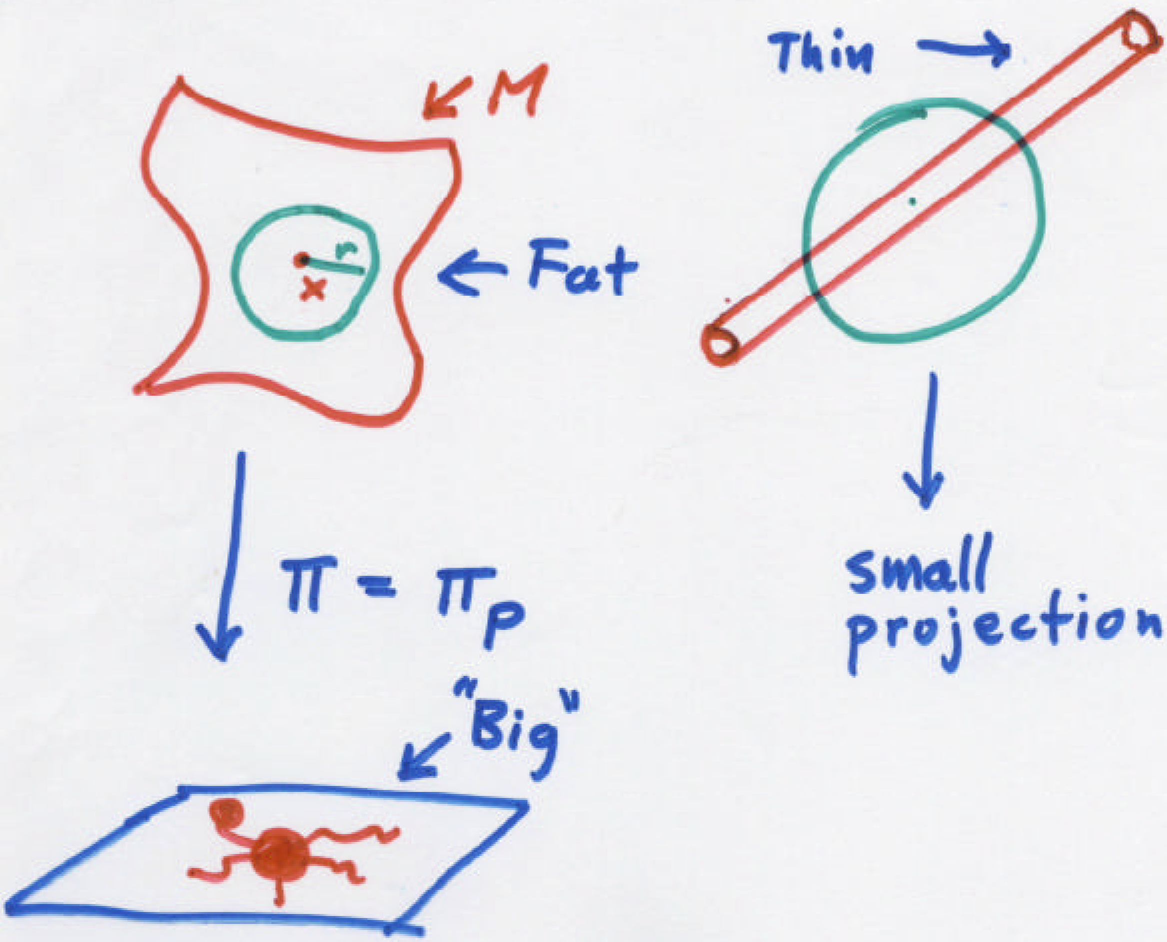
- The theorem can be formulated for an infinite set with a probability measure.
- $c_1 = c_0 \sqrt{n}$, where c_0 is a universal constant. Worst case scenario: $c_2 = c_0\, n$. If there is "roughly" a D-dimenional set carrying K, then $c_2 = c_0\, D$.
- Empirical bounds are better.

"Fat" d sets

M is "fat" for dimension d
if for all $x \in M$
$\exists$ a (d) hyperplane P
such that

$$\pi(M \cap B(x,r))$$

contains a Ball $B(y, \varepsilon r)$
on P



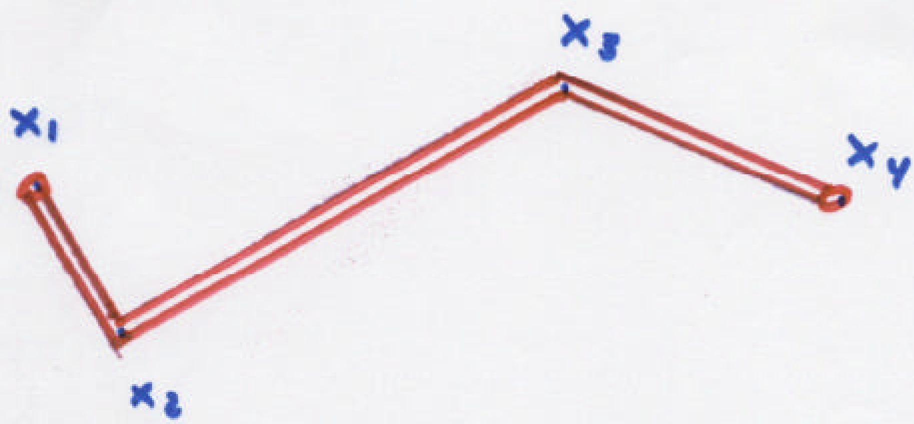
Why Demand Fat?

Let $\{x_1, \dots, x_N\} \in \mathbb{R}^n$,

$2 \leq d \leq n-1$

$\exists$ a diffeo S^d $(= M)$ containg all points,

$\text{Area}_d(M) < \varepsilon$



Thin hairs attached to surfaces are the bane of GMT.

Th. (J, Lerman to appear)
Lerman CPAM for $d = 1$

μ on $[0,1]^n$

$1 \leq d \leq n-1$

$\beta = \beta_2 = \beta_{2,d}$

Approximate by d Planes

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

$\Rightarrow$ $\exists$ "Good" d set M s.t.

(1) $\mu(M) \geq A^{-1}$

(2) $\text{Area}_d(M) \leq C_2 A$

Comments: Best Possible (except c_1, C_2 !!)
"Good" = Fat "Glueing" of Lipschitz Graphs (+ regularity)

Given: A measure $\mu \in P_r$ or points $\{x_j\} \equiv K \subseteq \mathbb{R}^n$ (Here set $\mu = \frac{1}{\#K} \sum \delta_{x_j}$)

Sample Problem: Which pieces of μ (points in K) are "one dimensional", "two dimensional", etc?

Answers: Find "short" curve (or "small" surface, etc.) hitting μ in large measure.

Philosophy (for Answer): points in the "short" curve are "one dimensional"

DESCRIPTION OF THE ALGORITHM

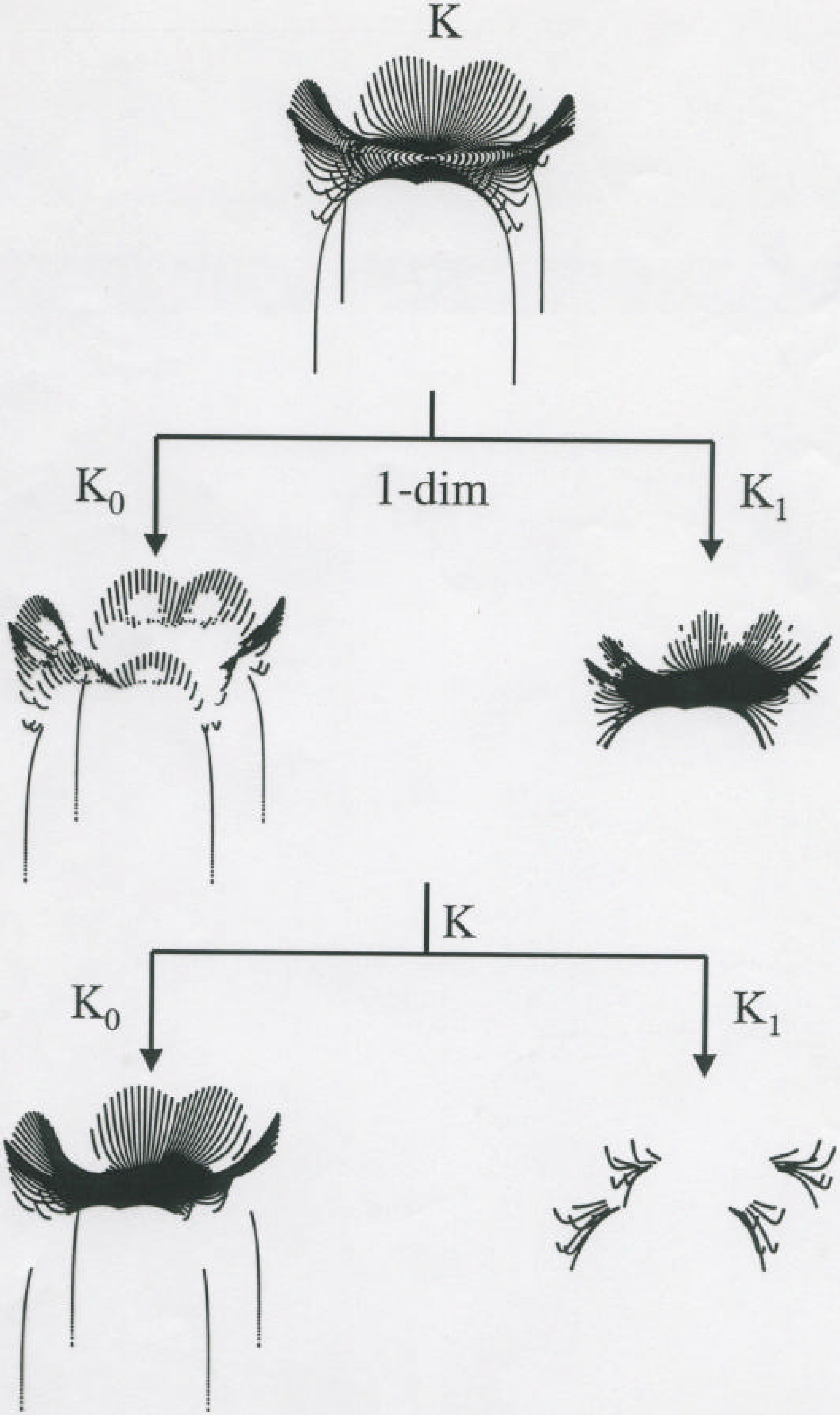
1. If K has structure - stop at K.

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

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

4. Let $$K_0 \equiv \{ x \mid m \leq J(x) < \tfrac{m+M}{2} \}$$ $$K_1 \equiv \{ x \mid \tfrac{m+M}{2} \leq J(x) \leq M \}$$

5. Repeat steps $1-5$ for $K = K_0$ and $K = K_1$.


K
K_0
1-dim
K_1
K
K_0
K_1

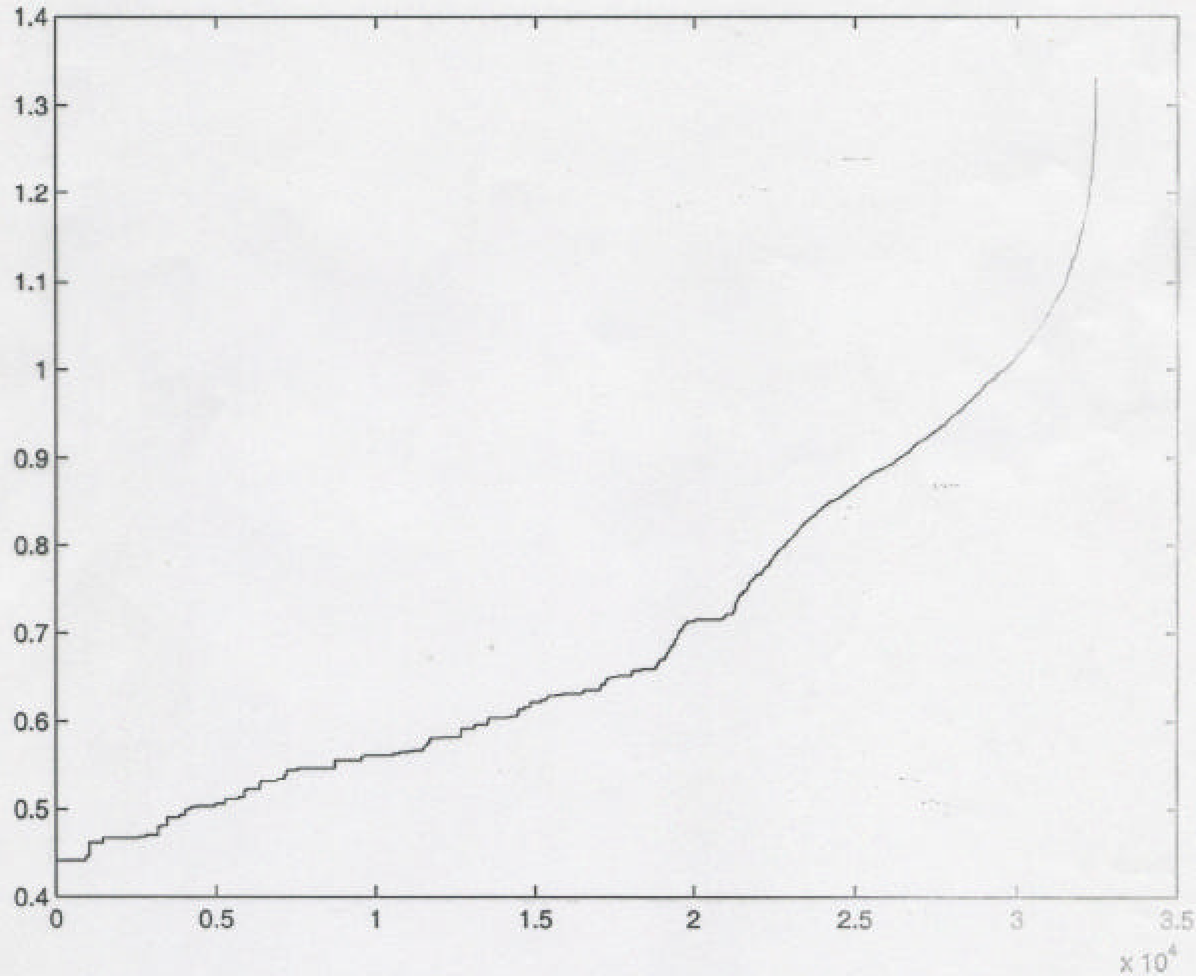
71

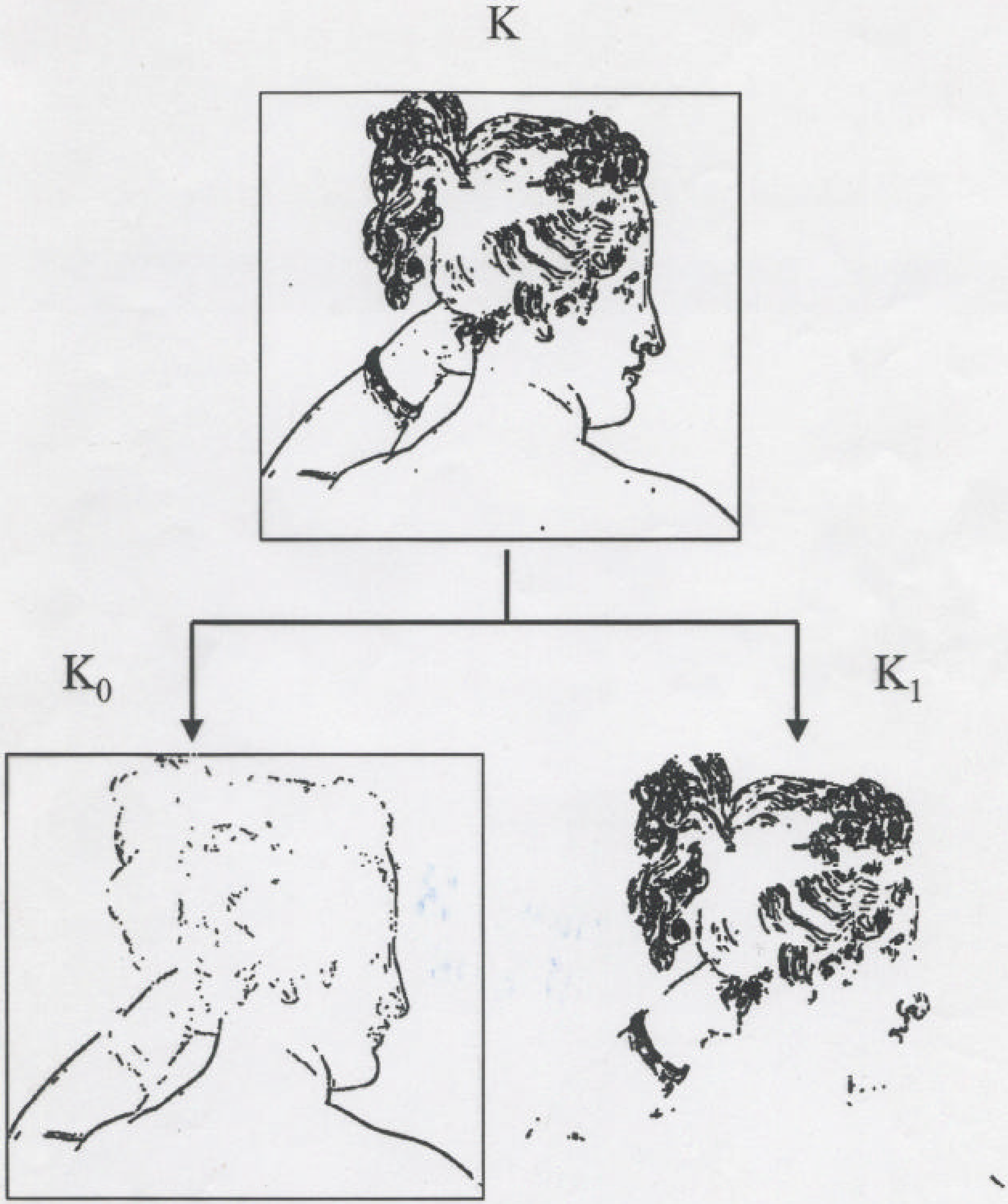
EXAMPLES OF SETS AND ANALYSES

EXAMPLE 1

Input: $\approx 80,000$ points in the plane given by the edge map of the image "Paolina".

Output: Partition into subsets with different geometric structures.

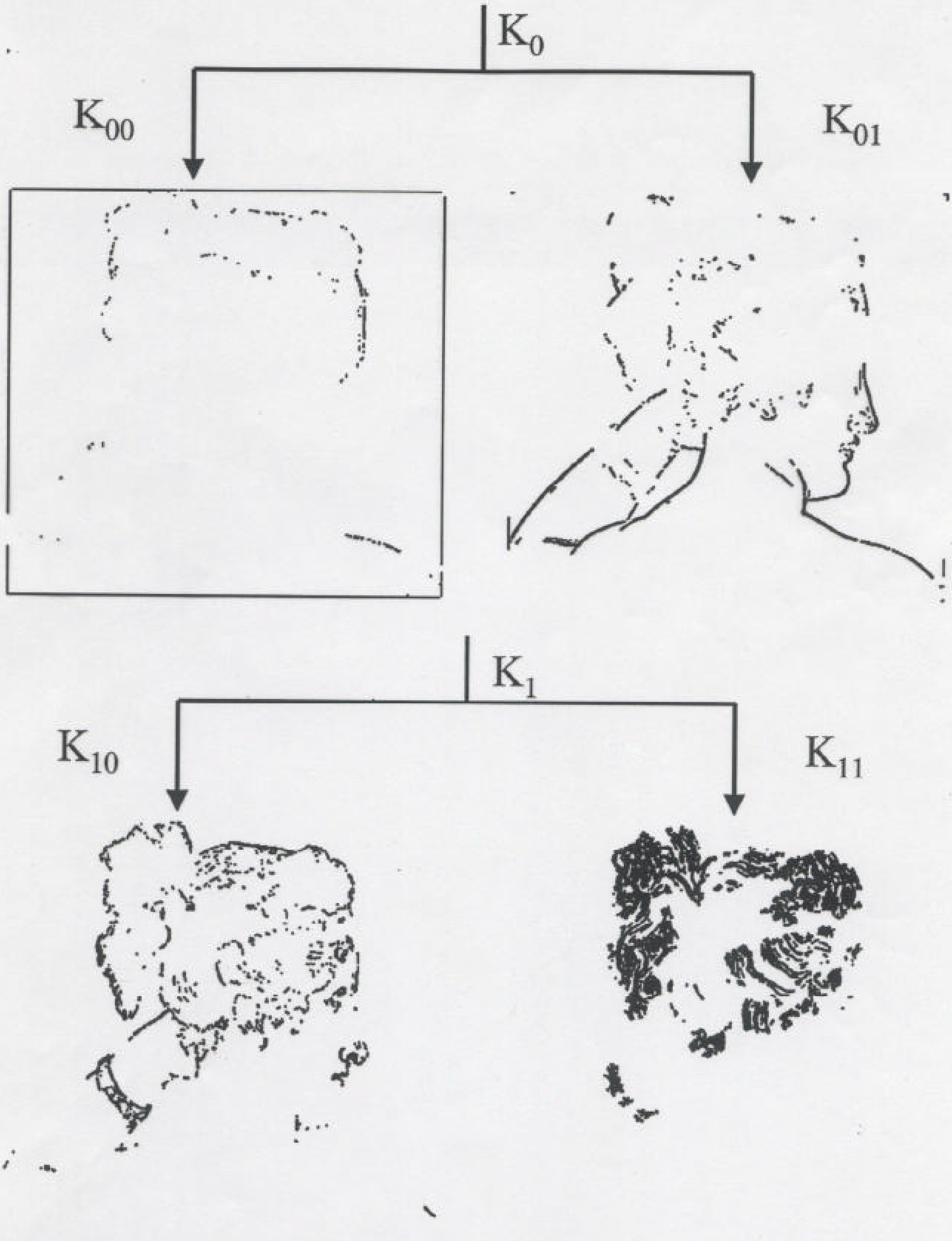

1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0
0.5
1
1.5
2
2.5
3
3.5
x 10^4

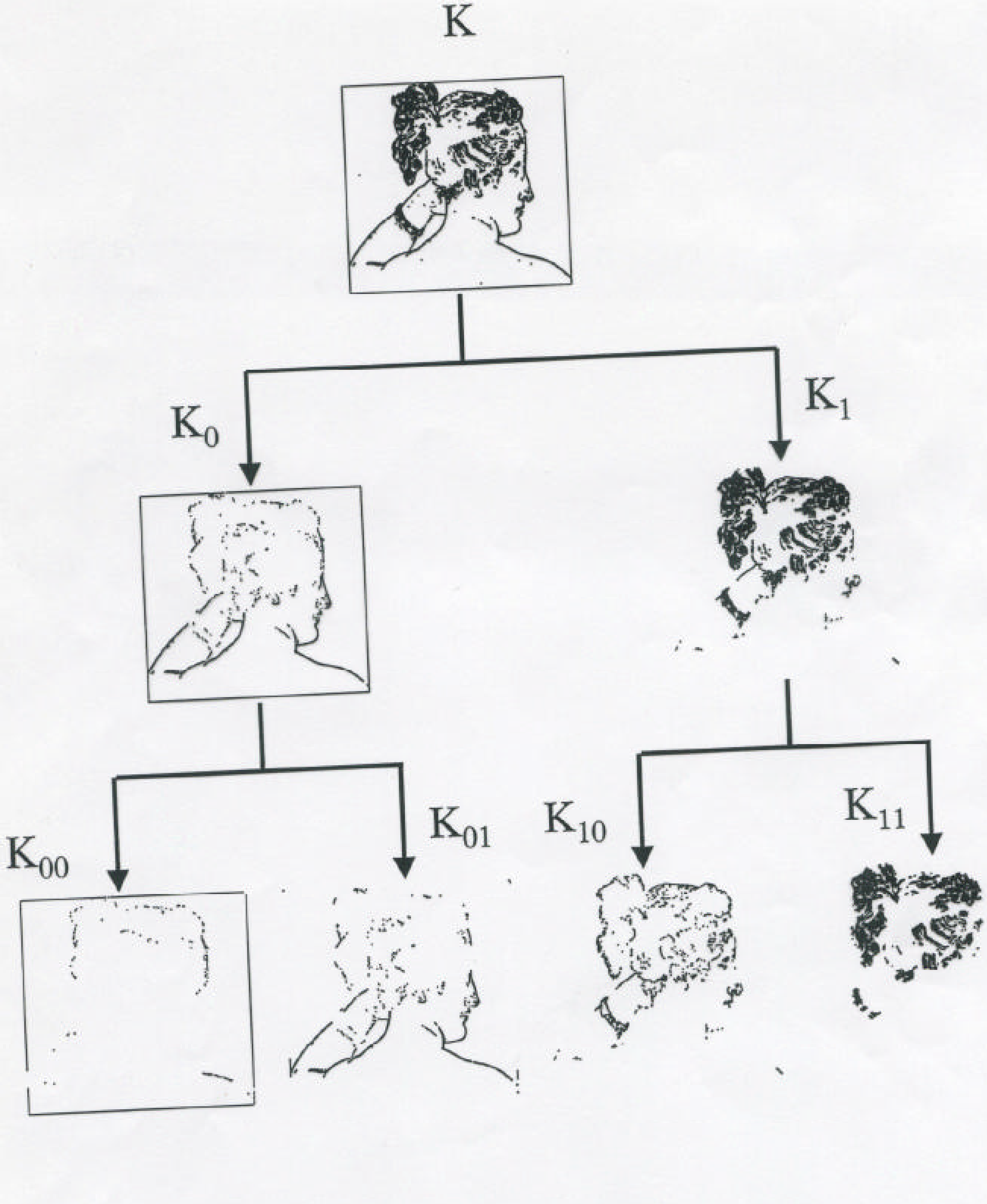


$$\sum_{Q \ni x} \beta^2(Q) \le M_1$$

74

- The new sets are partitioned again in the same way.


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


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

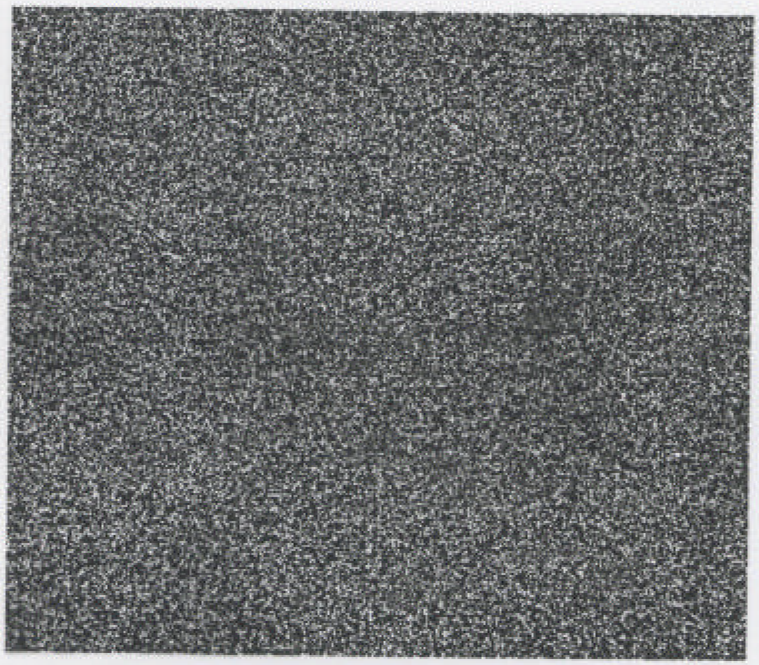
The Airplane Images

512*511 original image (jpeg, NASA AILS):

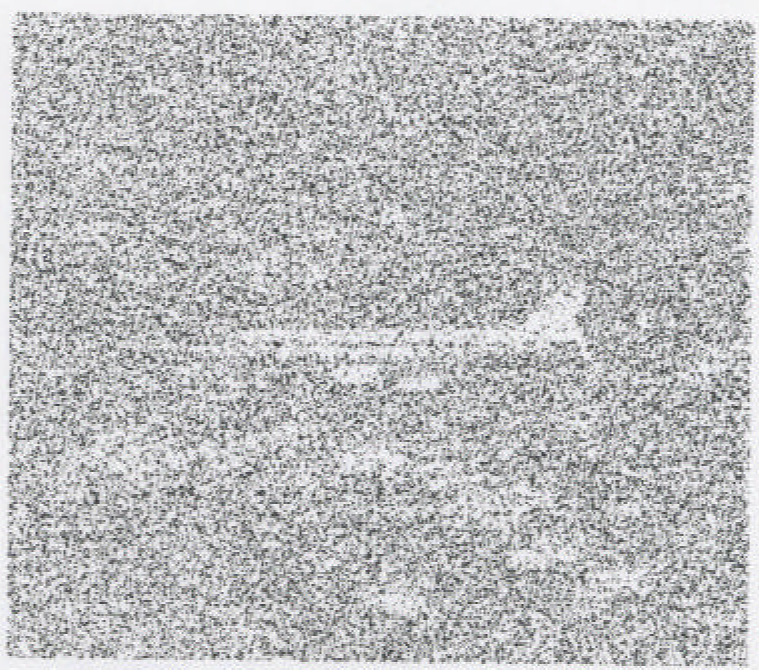
same image + noise ~ n(0,100):

Noisy image partition:

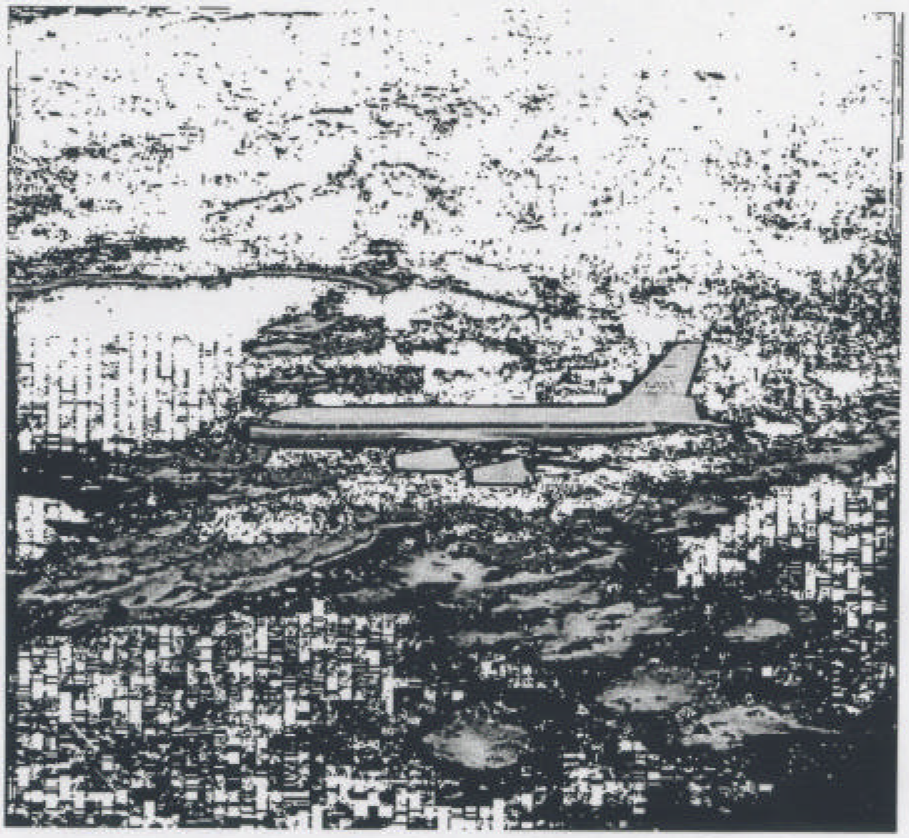
K_0



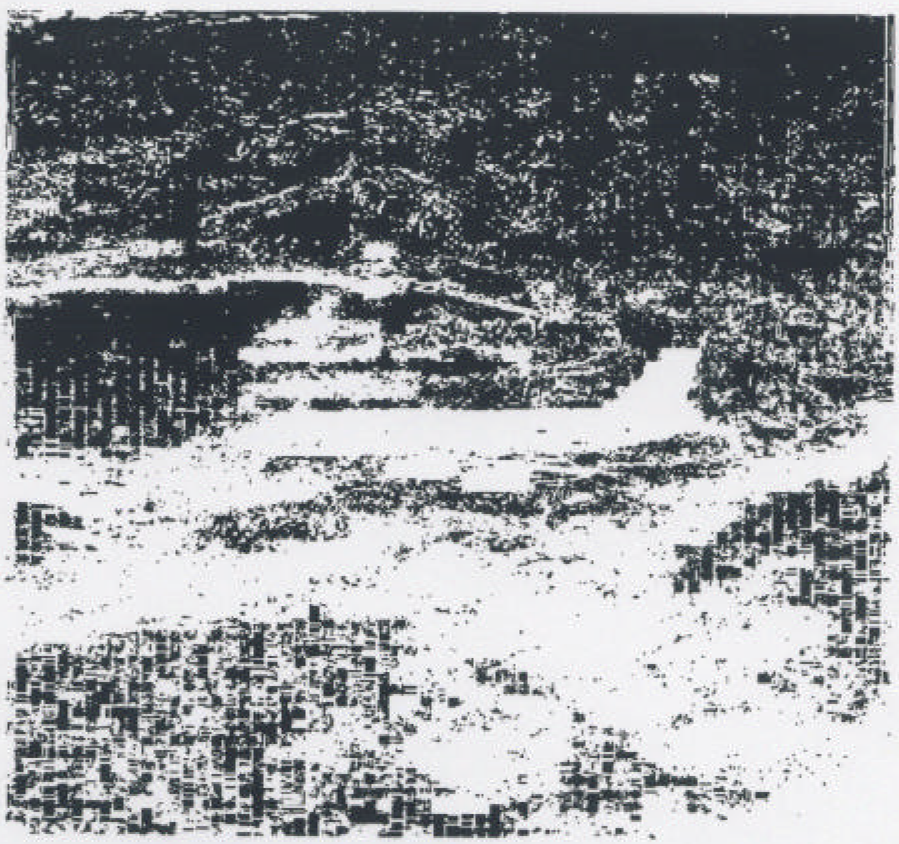
K_1

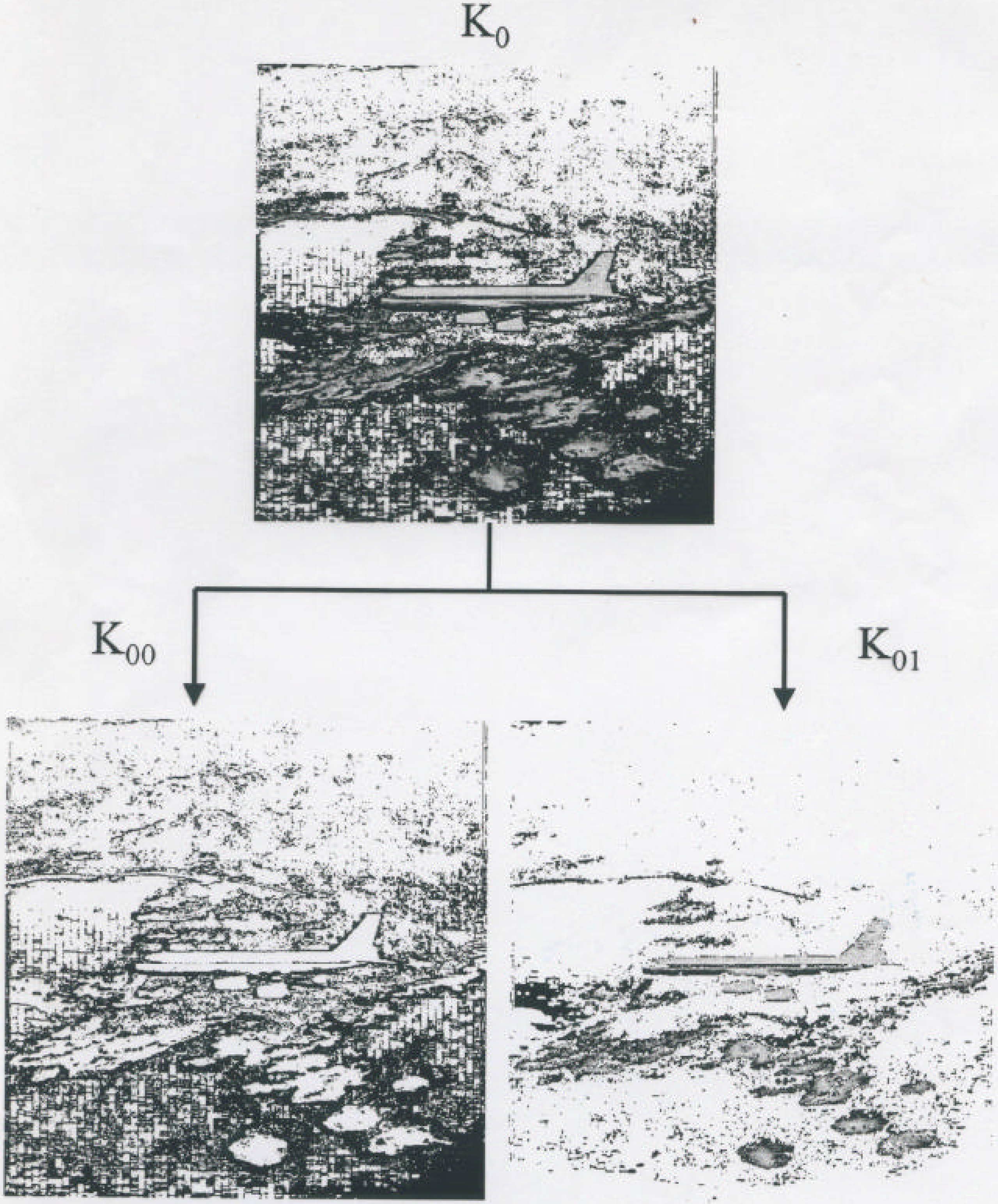


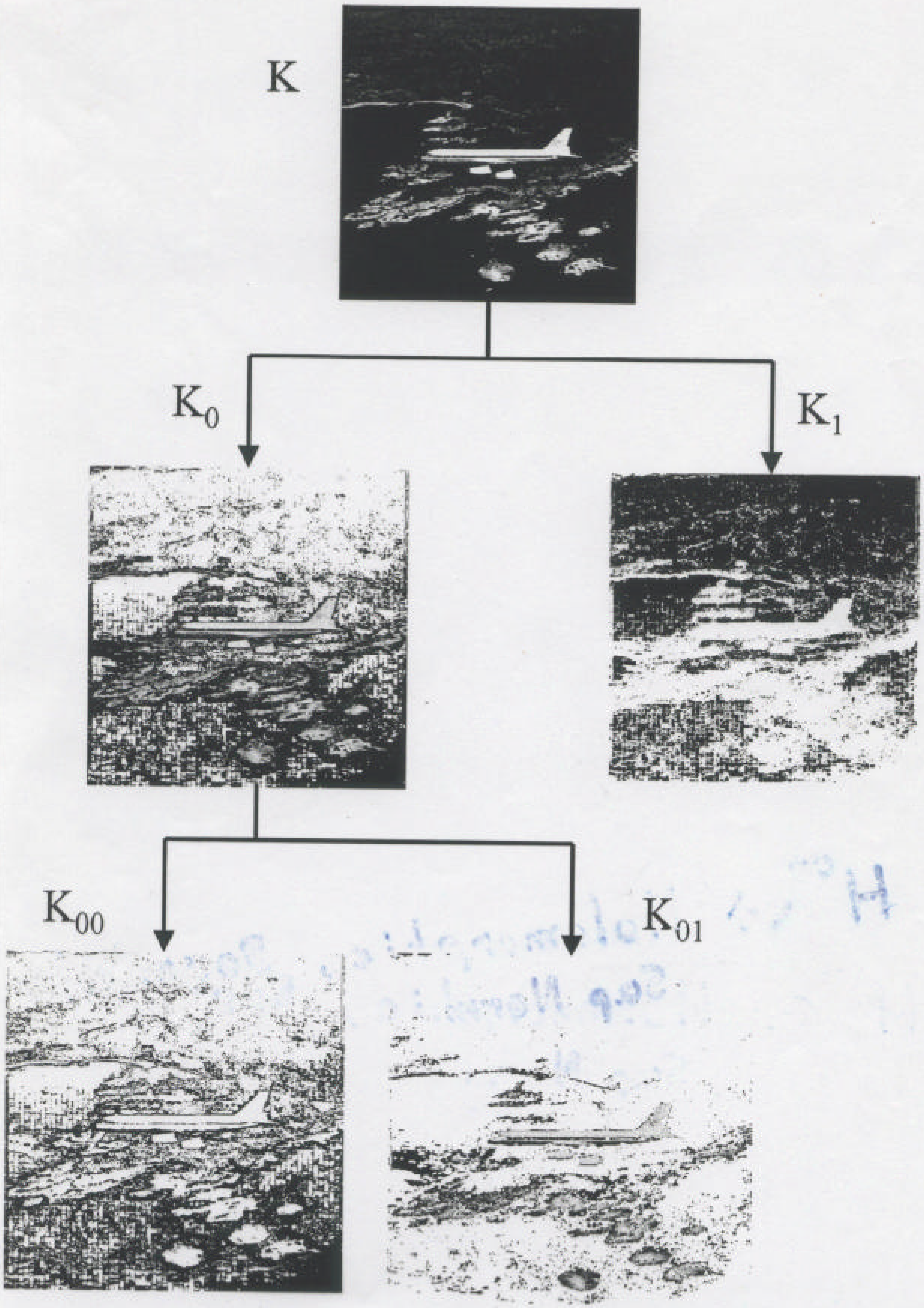
K_0



K_1




K_0
K_{00}
K_{01}


K
K_0
K_1
K_{00}
K_{01}