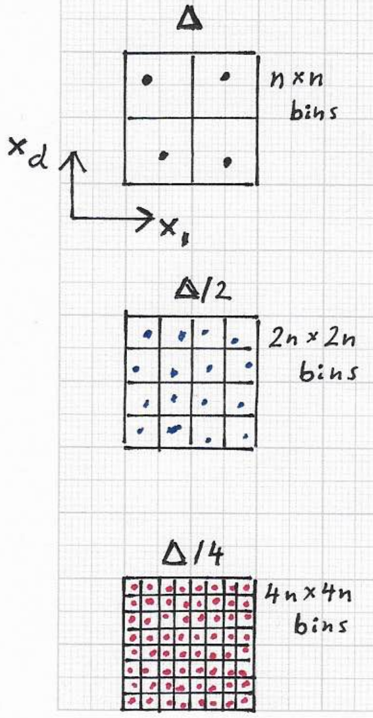
Stratovan

RICHARDSON EXTRAPOLATION FOR FEATURE SPACE – Cont'd.

- **SPARSENESS:** A high-dimensional feature space is sparsely populated with feature points/vectors of samples. It is generally desirable to (i) keep the dimensionality of feature space low; (ii) use the largest possible number of feature points/vectors; and (iii) use a small number of 'bins' when discretizing the intervals of considered feature axes/directions. Further, in order to perform Richardson extrapolation operations, one must have several feature point sub-sets with (ideally) geometrically decreasing point spacing.

- **Ex.:** Consider the idealized case of samples in a feature space that is discretized by (hyper-)cubes of geometrically decreasing spacing.



- Parameters:
  - d – number of dimensions
  - n – number of 'bins' (same for each dimension)
  - Δ, $\Delta/2$, $\Delta/4$, ... – spacing, geometrically decreasing

- Specific example: $d=5$, $n=4$:
  - resolution 0: $4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 4^5 = 1024$ bins
  - resolution 1: $8 \cdot 8 \cdot 8 \cdot 8 \cdot 8 = 8^5 = 32768$ bins
  - resolution 2: $16 \cdot 16 \cdot 16 \cdot 16 \cdot 16 = 16^5 = 1048576$ bins
  - **RESOLUTION k: $2^{kd} \cdot n^d$ bins**

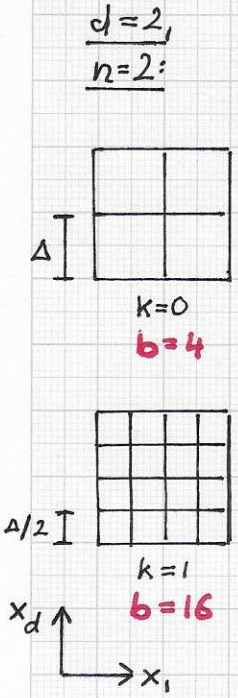
$\Rightarrow$ Sparsely sampled feature space must be used effectively, e.g., via stochastic sub-sampling.

BH

■ RICHARDSON EXTRAPOLATION FOR FEATURE SPACE – Cont'd.

• **Formula:** Given a d-dimensional feature space and using an initial resolution of n value 'bins' for each of the d dimensions, the number of (hyper-) cube 'bins' after k binary bin-subdivision steps is

$$b = 2^{kd} \cdot n^d .$$



[Related work by Ronald COIFMAN et al. concerns approaches for dealing with the sparseness of data in high-dim. feature space via: diffusion maps; geometric harmonics; embedding feature data in LOWER-DIM. EUCLIDEAN SPACE.]

• **Issues:** (i) For simplicity, one can assume that the Δ-value is the same for all dimensions, $\Delta = (max - min)/n$. (The values of max and min are maximal and minimal values defining the range of x-values for which feature data exist.)

(ii) "Standard" Richardson extrapolation assumes that the spacing decreases geometrically via halfing: Δ, $\Delta/2$, $\Delta/4$, ..., i.e., $\Delta_k = \Delta/2^k$, where $k = 0, 1, 2, \ldots$ denotes resolution level.

(iii) One must define the relationship between spacing $\Delta_k = \dfrac{max - min}{n \cdot 2^k}$ and $b_k = 2^{kd} \cdot n^d$.

RICHARDSON EXTRAPOLATION FOR FEATURE SPACE - Cont'd.

- Issues: ... Spacing and number of bins:

$$b_k = \alpha \cdot \Delta_k \Rightarrow \text{factor } \alpha = b_k / \Delta_k$$

$$\alpha = \left(2^{kd} \cdot n^d\right) / \left((max - min)/2^k n\right)$$

- Ex.: $(max - min) = 1$; $n = 1$:

resolution k	spacing Δ_k	b_k $d=1$	$d=2$	$d=3$	... $d=10$
0	1/1	1	1	1	1
1	1/2	2	4	8	1024
2	1/4	4	16	64	1048576

→ Use: If one needs to ensure a geometric progression of spacing values Δ_k — more specifically a halfing-based progression, $1, \frac{1}{2}, \frac{1}{4}, \ldots$ — then one will have to use a bin-number progression that satisfies the respective factor α.

- Ex.: $(max - min) = 1$; $d = 5$ (5-dim. feature space); $n = 4$ ("4 univariate bins" per dimension):
  - resolution 0: $\Delta_0 = 1/4$, $b_0 = 4^5 = 1024$
  - resolution 1: $\Delta_1 = 1/8$, $b_1 = 2^5 \cdot 4^5 = 32768$
  - resolution 2: $\Delta_2 = 1/16$, $b_2 = 2^{10} \cdot 4^5 = 1048576$

- Ex.: GIVEN IS A MAXIMAL NUMBER OF FEATURE POINTS, N. $(max - min) = 1$; $d = 5$. How large can n be when using all N points? (On average, at least 1 point per bin is desired.)

$$\overset{k=2}{\Rightarrow} b_2 = 2^{2 \cdot 5} \cdot n^5 = 2^{10} \cdot n^5 \leq N \Rightarrow n^5 \leq N/2^{10} \Rightarrow n \leq \sqrt[5]{N/2^{10}}$$

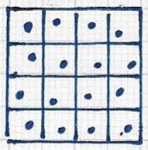
$$\Rightarrow \text{e.g., } N = 32768: n \leq 2$$

■ RICHARDSON EXTRAPOLATION FOR FEATURE SPACE – Cont'd.

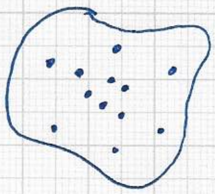
FEATURE SPACE

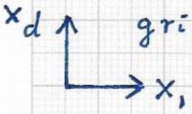
- Issues:… It is unreasonable to assume that the number of available feature value data in HIGH-DIM. feature space make possible a 'dense packing', e.g., on the typical 'hyper-hexahedral structure' of $b_{x_1} \times b_{x_2} \times \ldots \times b_{x_d} = b^d$ bins (hyper-cubes) in a d-dimensional feature space using n 1-dimensional bins per dimension $(x_1, x_2, \ldots, x_d)$.

d=2



hyper-cubes





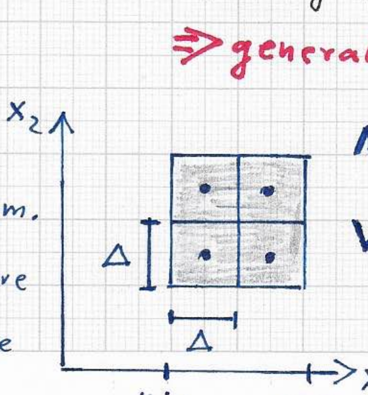
gridless

⇒ **Generally, one must give up the nice hyper-hexahedral structure / grid for high-dimensional feature space and use a GRIDLESS, SCATTERED DATA approach to define multiple resolutions with soundly defined spacing.**

(iv) One must define the relationships between the volume/size of the feature space region that contains feature data; the density of feature data in this region; and a 'meaningful' (average) spacing value for the 'data packing'.

⇒ generalize hyper-hexahedral structure to scattered data!

2-dim. feature space



total region volume : V

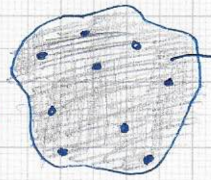
$N = 2 \cdot 2 = 4$ (= no. of samples)

$V = (2\Delta)^2 = 4\Delta^2$ (= total volume)

$\Rightarrow \Delta = \sqrt{V/4}$

more ⇒ general

$N = 9$



V = total volume

$\Rightarrow \Delta = \sqrt{V/N}$

"Formula for Δ assumes a spacing of hyper-hexahedra."

RICHARDSON EXTRAPOLATION FOR FEATURE SPACE – Cont'd.

... (iv) Spacing value computation for scattered feature data...

Ex.: 2-d space $\Rightarrow \Delta = \sqrt{V/N}$

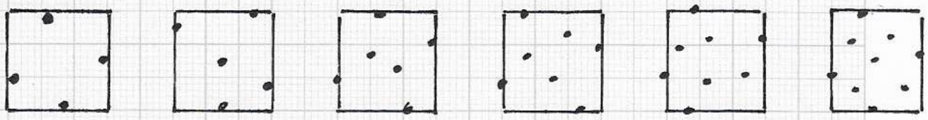
3-d " $\Rightarrow \Delta = \sqrt[3]{V/N}$

d-dim. " $= \Delta = \sqrt[d]{V/N}$

$\Rightarrow$ estimate of hyper-volume of region in d-dim. space containing N data: **minimal bounding box of N data has volume V.**

Ex.: $d = 2$, $V = 1$

$N=4$ $N=5$ $N=6$ $N=7$ $N=8$ $N=9$



$\Delta =$ $\sqrt{1/4}$ $\sqrt{1/5}$ $\sqrt{1/6}$ $\sqrt{1/7}$ $\sqrt{1/8}$ $\sqrt{1/9}$

(v) Hyper-volume V in d-dim. feature space $\rightarrow$ N feature points/vectors have (average) spacing $\Delta = \sqrt[d]{V/N}$.

$\Rightarrow$ Feature point sub-sets with k points have spacing

$$\Delta_k = \sqrt[d]{V/k} \; .$$

$\Rightarrow$ Richardson extrapolation can be done by using multiple spacing values Δ_k and extrapolating a needed quantity for $\Delta_\infty = 0$.

($\Delta_\infty = 0$: "infinitely dense sampling")

~BH