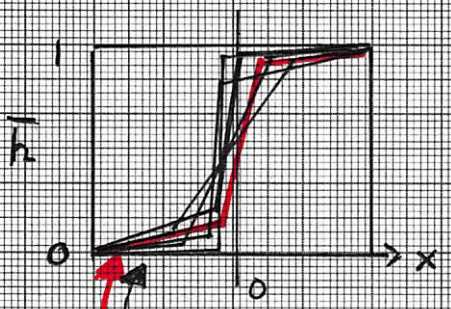
Stratovan

CUMULATIVE DISTRIBUTION FUNCTIONS – PCA, BEST APPROXIMATION, CLASSIFICATION



• black: cumulative distrib. fcts. of histograms of SAME MATERIAL (from different "segments")

• red: (piecewise linear representation of) MOST IMPORTANT "EIGEN-DISTRIBUTION" (with largest eigenvalue)

1) Apply PCA to cumulative distr. fcts. (with x-direction being the function direction), using a piecewise constant representation.

2) Using the piecewise constant representation of the $x(\bar{h})$ fcts, view each fct. as a point/vector in K-dim. space (and perform PCA for N such fcts.).

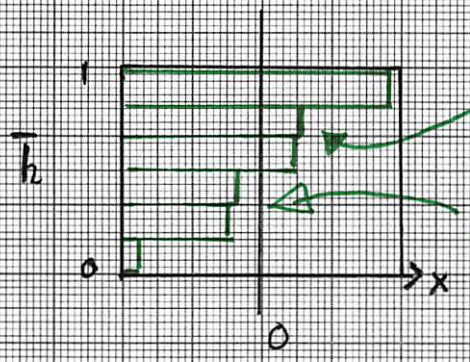
3) There are N $x(\bar{h})$ fcts., each approximated by K=N constant segments.

4) PCA produces K eigenvalues with associated eigenvectors – "EIGEN-DISTRIBUTIONS" – sorted based on absolute eigenvalue magnitude. THUS, the most important (or small set of most important) eigen-distribution(s) can be selected.

5) PROBLEM: Given a new cumulative distr. fct. of a newly scanned object/segment, does it represent an instance of THIS MATERIAL?

Stratovan

CUMULATIVE DISTRIBUTION FUNCTIONS – ... Cont'd.



K constant segments

green: piecewise constant approx. of $x(\bar{h})$ function of a newly scanned object/segment (approximating the cumulative distrib. fct. with K segments)

6) Compute best approximation of $x(\bar{h})$ using small set of most important eigen-distribution fcts. for the specific material [done ONLY for a newly scanned object/segment with a mean that is "very close" to the mean of this specific material]: COMPUTE BEST APPROXIMATION

$$a = \sum_{i=1}^{L} c_i \cdot e_i \quad ,$$

where the L most important eigen-distribution fcts. e_i, $i=1,\dots,L$, are used to detim least squares a:

mutually orthogonal

$\langle \cdot, \cdot \rangle$ inner product

$$\begin{pmatrix} \langle e_1, e_1 \rangle & \dots & \langle e_1, e_L \rangle \\ \vdots & & \vdots \\ \langle e_L, e_1 \rangle & \dots & \langle e_L, e_L \rangle \end{pmatrix} = \begin{pmatrix} \langle e_1, e_1 \rangle & & 0 \\ & \ddots & \\ 0 & & \langle e_L, e_L \rangle \end{pmatrix}$$

since $\langle e_i, e_j \rangle = 0$ for $i \neq j$

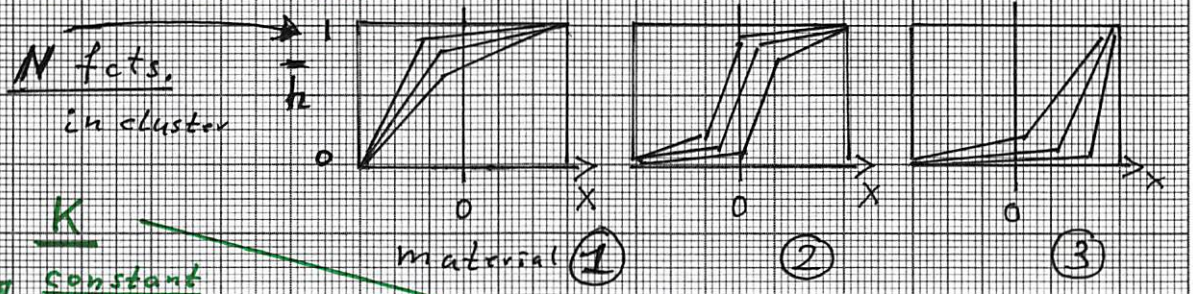
$$\Rightarrow c_i = \langle e_i, x \rangle$$

$\Rightarrow$ THE "SPECTRUM" of resulting $|c_i|$ values used to DETERMINE WHETHER $x(\bar{h})$ belongs to THIS MATERIAL.

CUMULATIVE DISTRIBUTION FUNCTIONS – CONT'D.

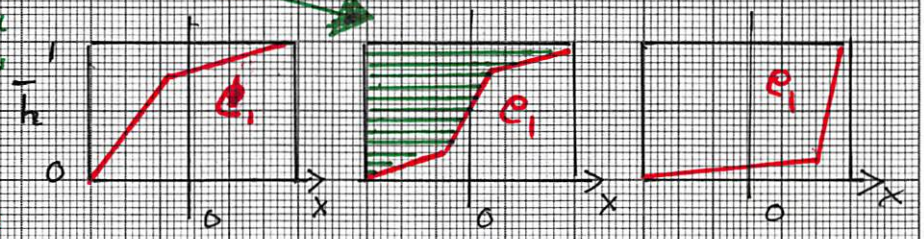
- **IDEA:** If $x(\bar{h})$ can be "very well approximated by a, i.e. $\|x(\bar{h}) - a\| \ll 1$, then $x(\bar{h})$ is well approximated by a small number of (principal) eigen-distributions and $x(\bar{h})$ therefore represents another instance of this specific material.

- **Example:** 3 cumulative distrib. fct. clusters for 3 materials $\Rightarrow$ 3 clusters:



K constant segments used to perform all computations on K-dim vectors

$\Rightarrow$ Use K=N



most important eigen-distribution fcts. for materials (1), (2), (3)

7) CLASSIFICATION: $x(\bar{h})$ approximated well by expansion a using eigen-distrib. fcts. of material $M \Leftrightarrow x(\bar{h})$ represents M.

- **NOTE:** Using INDEPENDENT COMPONENT ANALYSIS (ICA) could be used as well.

BH