Bernd Hamann — 1 — 11/04/2019

Stratovan

■ EIGENFUNCTIONS AND CLASSIFICATION

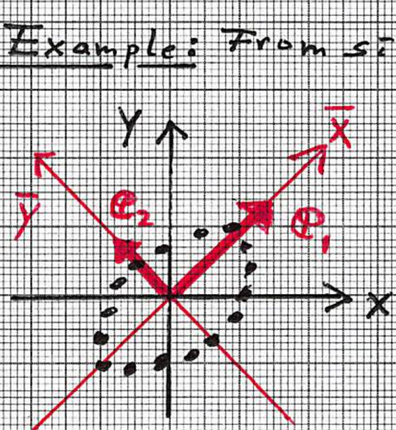
• **Idea:** → "Cocktail Party Problem": Multiple sound sources produce a single signal that must be separated into its constituents...

→ Principal Component Analysis (PCA) and Independent Component Analysis (ICA) can be used to determine "components" of point data sets or function sets.

→ Signatures (signature functions), e.g., histograms, of materials in scans can be viewed as functions whose underlying components or eigenfunctions can be computed.

→ **GOAL:** Identify the underlying components of a recorded/scanned/derived signature.

• **Example:** From simple 2D PCA to functions:



$\{e_i\}$ orthogonal basis

given: point set $\{\bullet\} = \{p_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}\}_{i=1}^{n}$ (after mean subtraction)

computed: eigenvalue-scaled eigenvectors e_1 and e_2 from covariance matrix:

$$\begin{bmatrix} x_1 & \dots & x_n \\ y_1 & \dots & y_n \end{bmatrix} \cdot \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} = \begin{bmatrix} \sum x_i^2 & \sum x_i y_i \\ \sum x_i y_i & \sum y_i^2 \end{bmatrix}$$

▪ EIGENFUNCTIONS & HISTOGRAM DISTANCES

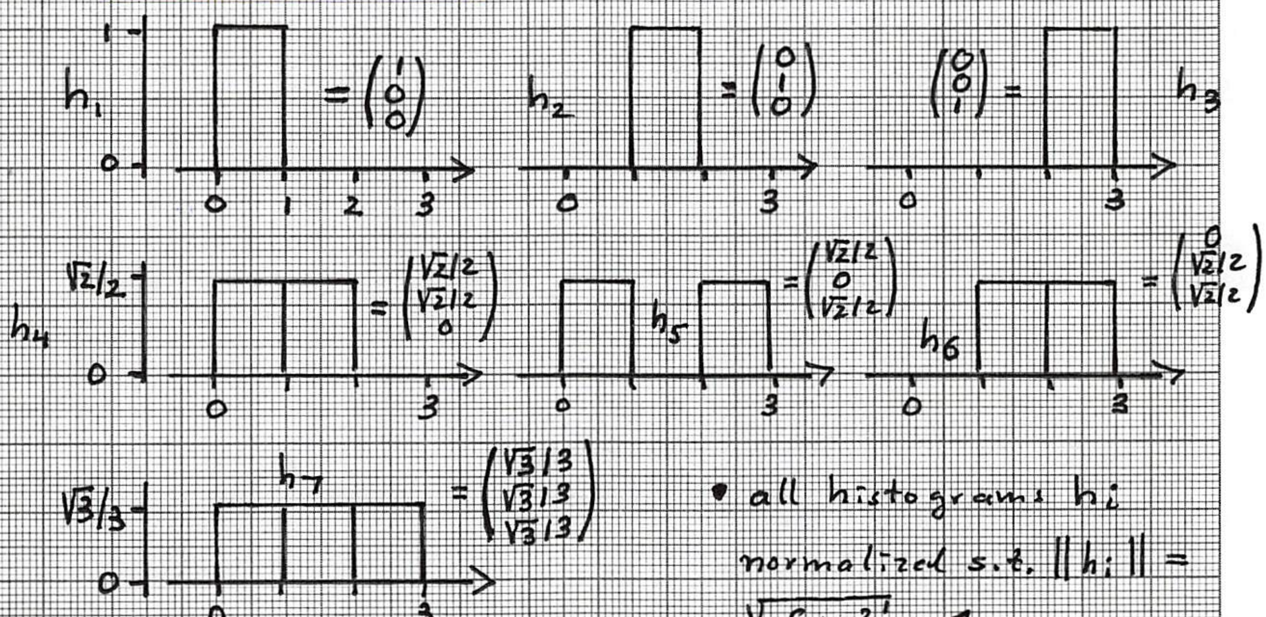
GENERAL PROBLEM: → Need for better distance measure!

→ Interpreting a histogram with N bins as a vector in N-dim. space is **NOT APPROPRIATE** when using an "angle-between-vectors" as distance measure!

Example:
- 7 histograms with 3 bins
- Histogram viewed as a piecewise constant function f
- f normalized such that

$$\|f\| = \sqrt{\int f^2} = 1$$

• **WANTED:**
$\text{dist}(h_1, h_2) < \text{dist}(h_1, h_3)$



$h_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$, $h_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}$, $\begin{pmatrix}0\\0\\1\end{pmatrix} = h_3$

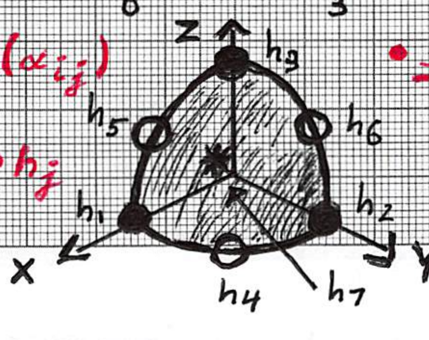
$h_4 = \begin{pmatrix}\sqrt{2}/2\\\sqrt{2}/2\\0\end{pmatrix}$, $h_5 = \begin{pmatrix}\sqrt{2}/2\\0\\\sqrt{2}/2\end{pmatrix}$, $h_6 = \begin{pmatrix}0\\\sqrt{2}/2\\\sqrt{2}/2\end{pmatrix}$

$h_7 = \begin{pmatrix}\sqrt{3}/3\\\sqrt{3}/3\\\sqrt{3}/3\end{pmatrix}$

- all histograms h_i normalized s.t. $\|h_i\| = \sqrt{\int h_i^2} = 1$
- Interpret h_i as vector $\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix}$
- Vectors $\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix}$ are positional vectors of point on unit sphere.

$\cos(\alpha_{ij}) = h_i \bullet h_j = c_{ij}$

dot product



• **BAD:** $c_{12} = c_{13} = c_{23} = 0$

⇒ **BETTER DISTANCE MEASURE NECESSARY !!!**

■ HISTOGRAM DISTANCES

• Reference: "Geodesic PCA in the WASSERSTEIN Space" by Bigot, Gouet, Klein & López

→ HISTOGRAM DISTANCE MEASURE MUST CONSIDER "DISTANCE OF HISTOGRAMS IN X-COORDINATE / ABSCISSA DIRECTION"...

→ Example from previous page:

- $dist(h_1, h_2) = dist(h_2, h_3)$
- $dist(h_1, h_3) > dist(h_1, h_2)$ ...

{→ Note: i) See previous page: "dot product metric" not appropriate

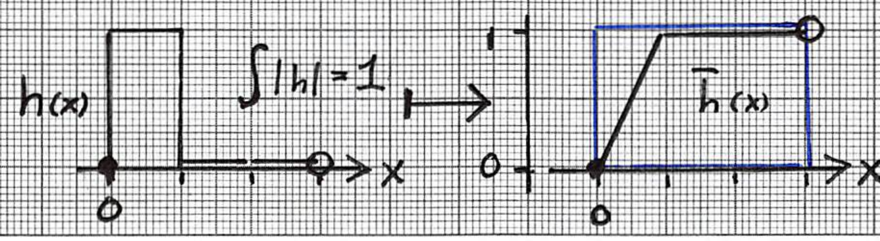
ii) Metric $dist(h_i, h_j) = \sqrt{\int (h_i - h_j)^2}$ not appropriate (considering y-coordinate / ordinate direction only) }

→ WASSERSTEIN SPACE: CONSIDERS X-DIRECTION DISTANCE!

(See: "Optimal Mass Transport"... by Kolouri & Soheil et al., IEEE Signal Processing..., 2017)

"Earth Mover's Distance": special Wasserstein distance

→ CUMULATIVE DISTRIBUTION essential for x-direction distance!



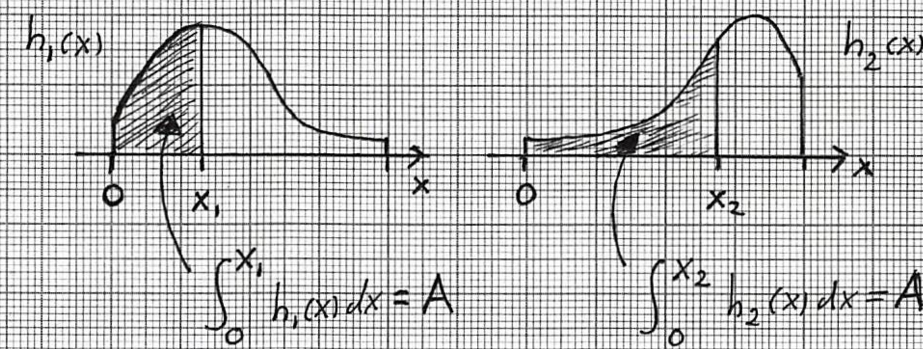
cumulative distribution: $\bar{h}(x) = \int_0^x h(\tilde{x})\, d\tilde{x}$

Stratovan

■ HISTOGRAM DISTANCES

- **Principle: "Earth Mover's Distance"**

  → given: distribution $h_1(x)$ and $h_2(x)$,
  with $\int h_1(x) = \int h_2(x) = 1$



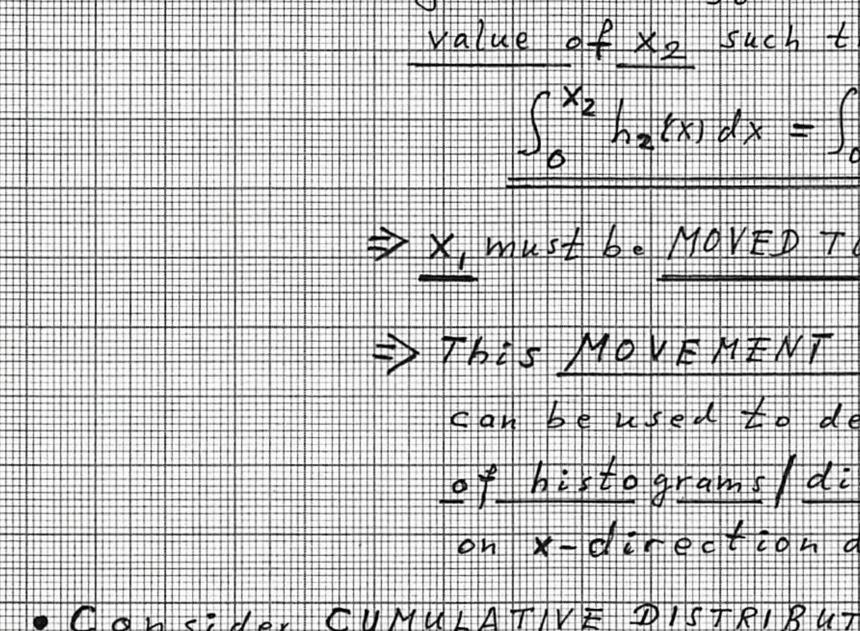
⇒ given x_1 and $\int_0^{x_1} h_1(x)\,dx$, what is the value of x_2 such that

$$\int_0^{x_2} h_2(x)\,dx = \int_0^{x_1} h_1(x)\,dx = A. \quad (*)$$

⇒ x_1 must be MOVED TO x_2 to satisfy (*)

⇒ This MOVEMENT IN x-DIRECTION can be used to define a DISTANCE of histograms/distributions based on x-direction distance!

- Consider CUMULATIVE DISTRIBUTION: $\overline{h}_1(x), \overline{h}_2(x)$



- **Principle:**
  - Cumulative values are the same for corresponding pairs (x_i^1, x_i^2).
  - Possible total movement cost: $\sum_i |x_i^1 - x_i^2|$

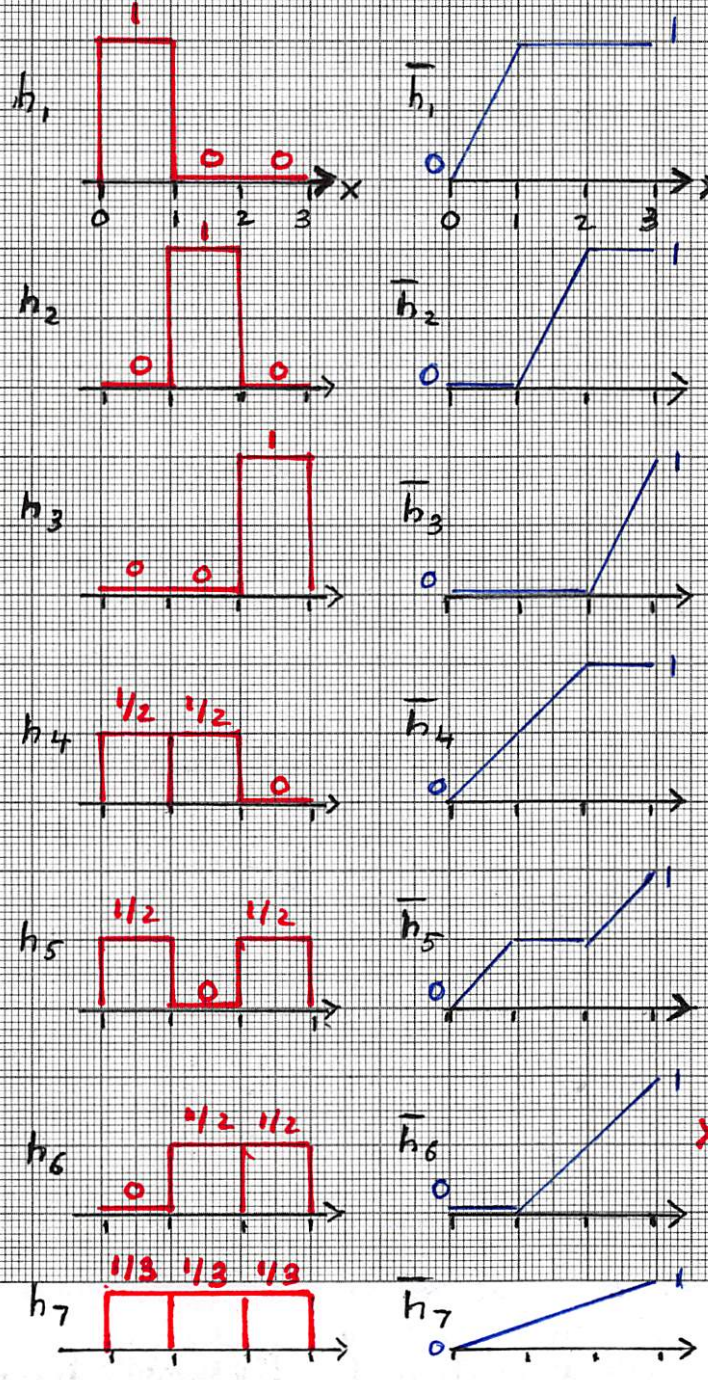
- Cumulative distributions should be STRICTLY MONOTONICALLY INCREASING.

Stratovan

HISTOGRAM DISTANCES

- **IDEA:** DEFINE DISTANCES BASED ON DISTANCES OF CUMULATIVE DISTRIBUTIONS IN X-DIRECTION.

- **Example:** Distributions given as histograms (= piecewise-constant functions)



$h_i(x)$ given histogram

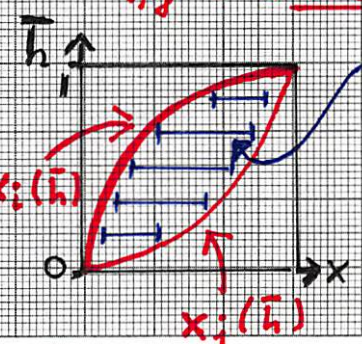
$\bar{h}_i(x)$ cumulative function (consisting of linear & constant segments)

- **Define WASSERSTEIN distance for $\bar{h}_i$ and $\bar{h}_j$:**

$$\boxed{d_{i,j}} = \int_0^1 \left| \bar{h}_j^{(-1)}(x) - \bar{h}_i^{(-1)}(x) \right| d\bar{h}$$

$$\boxed{= \int_0^1 \left| x_j(\bar{h}) - x_i(\bar{h}) \right| d\bar{h}}$$

- $d_{i,j}$ uses x-direction distance:



$|x_j(\bar{h}) - x_i(\bar{h})|$

- $d_{i,j}$ integrates absolute values of differences of $x_j(\bar{h})$ and $x_i(\bar{h})$.

■ HISTOGRAM DISTANCES

• Example: Compute distances for the seven cumulative functions $\bar{h}_1, \ldots, \bar{h}_7$:

$d_{i,j}$	$\bar{h}_1$	$\bar{h}_2$	$\bar{h}_3$	$\bar{h}_4$	$\bar{h}_5$	$\bar{h}_6$	$\bar{h}_7$
$\bar{h}_1$	0						
$\bar{h}_2$	1	0					
$\bar{h}_3$	2	1	0				
$\bar{h}_4$	1/2	1/2	3/2	0			
$\bar{h}_5$	1	3/4	1	1/2	0		
$\bar{h}_6$	3/2	1/2	1/2	1	1/2	0	
$\bar{h}_7$	1	1/4	1	1/2	1/4	1/2	0

NOTE: $d_{1,2} = 1$ and $d_{1,3} = 2$ AS DESIRED !!!

! THUS: The $d_{i,j}$ values define desired distance values for cumulative (distribution) functions and thus meaningful distances for histograms $h_i(x)$ and $h_j(x)$ – by determining x-direction distances. !

• Relevance: Wasserstein distance supports
(i) more meaningful HISTOGRAM DISTANCES;
(ii) better HISTOGRAM CLUSTERING; and
(iii) improved HISTOGRAM AVERAGING.

Bernd Hamann

11/21/2019
Stratovan

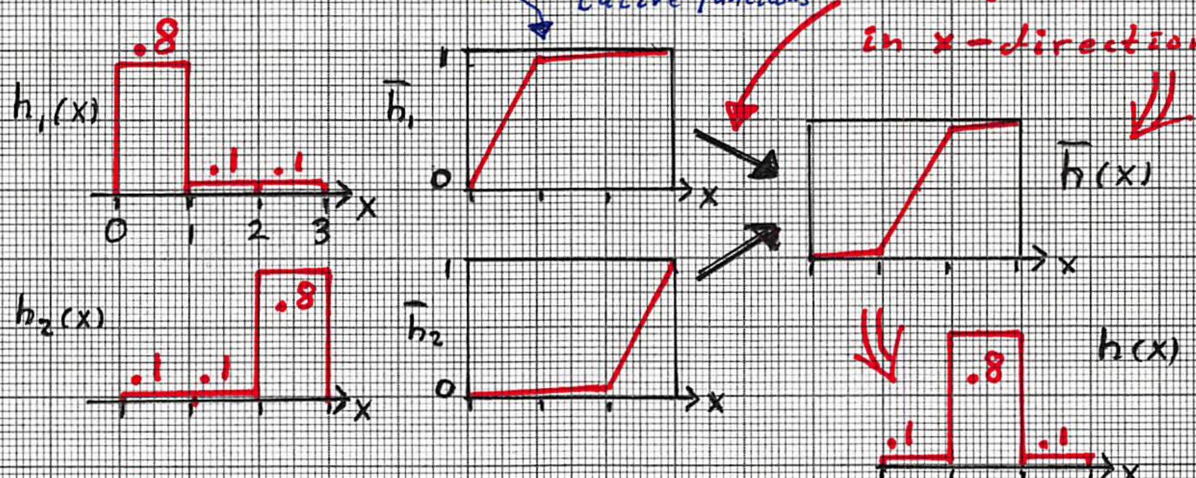
▪ HISTOGRAM DISTANCES — AVERAGING IN WASSERSTEIN SPACE

• **IDEA:** Given two histograms, first compute their cumulative (distribution) functions; second, compute the average of the two cumulative functions in "Wasserstein space" (by averaging in x-direction); third, determine a histogram based on the computed average in Wasserstein space.

• Examples:

must handle "flat" segments in cumulative functions

average $\bar{h}_1$ and $\bar{h}_2$ plots in x-direction



• $\frac{1}{2}(h_1(x)+h_2(x))$ is NOT a proper average of the two histograms!

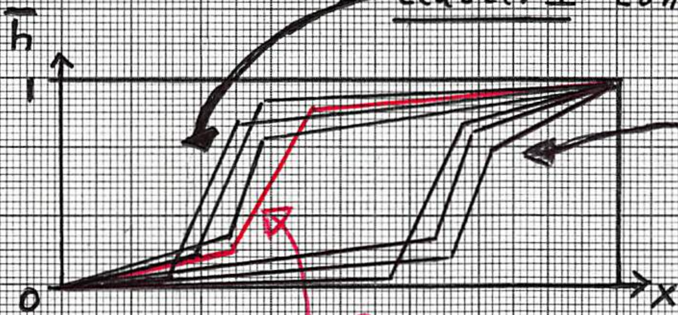
• $h(x)$ is a desirable average of $h_1(x)$ and $h_2(x)$!
$\Rightarrow$ Wasserstein space based averages are are proper averages, e.g., for histogram cluster center computation and other averaging-based characteristics of histograms.

HISTOGRAM DISTANCES – CLUSTERING AND CLASSIFICATION IN WASSERSTEIN SPACE

- **IDEA:** Based on the "x-direction distance measure of cumulative histograms" ("Wasserstein space distance"), define CLUSTERS of same material class. Compute CLUSTER MEANS also in this space. Compute the x-direction distances between a (new) material to be classified and cluster means. Computed distance values to determine whether the (new) material can be associated with any of the material clusters – subject to satisfying a distance threshold.

- Example:

cluster 1: consisting of functions $x_1(\bar{h})$, $x_2(\bar{h})$, $x_3(\bar{h})$



cluster 2: consisting of functions $x_4(\bar{h})$, $x_5(\bar{h})$, $x_6(\bar{h})$

RED: function $x(\bar{h})$ to be classified

→ Cluster means ("mean functions") of cluster 1 and cluster 2 are $X_1(\bar{h}) = \sum_{i=1}^{3} x_i(\bar{h}) / 3$ and

$$X_2(\bar{h}) = \sum_{i=4}^{6} x_i(\bar{h}) / 3 .$$

■ HISTOGRAM DISTANCES – CLASSIFICATION IN WASSERSTEIN SPACE

• **Example:** cont'd...

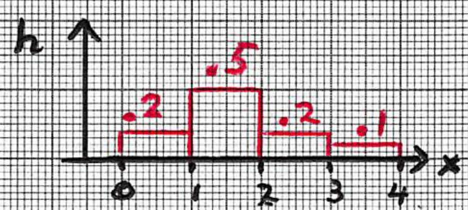
→ Compute distances between means $X_i(\bar{h})$ and to-be-classified function $x(\bar{h})$. Possibilities:

$$(i)\quad d(X_i, x) = \int_0^1 \left| X_i(\bar{h}) - x(\bar{h}) \right| d\bar{h}$$

$$(ii)\quad \hat{d}(X_i, x) = \left(\int_0^1 \left(X_i(\bar{h}) - x(\bar{h}) \right)^2 d\bar{h} \right)^{1/2}$$

⇒ IF d (or $\hat{d}$) $< \varepsilon$ THEN x "belongs to" the respective cluster(s).

→ Practical considerations:



→ original histogram (binned) $h(x)$ = piecewise CONSTANT function

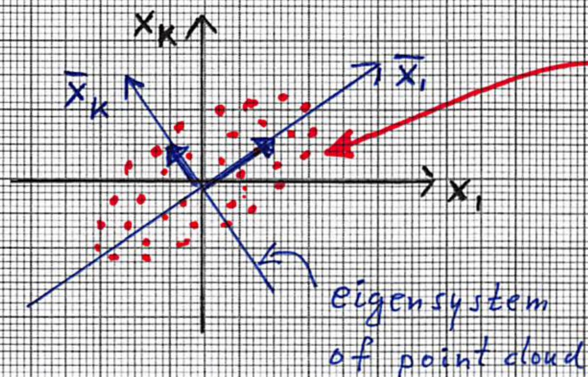
→ cumulative function $\bar{h}(x)$ = piecewise LINEAR function (with potentially zero-slope segments!) must be handled!

→ inverted function $x(\bar{h})$ is a piecewise LINEAR function.

~BH

■ HISTOGRAM DISTANCES – PCA, COVARIANE, "EIGENVECTORS FOR HISTOGRAM-DERIVED DATA AND FUNCTIONS"

1) PCA for point sets: (afer MEAN subtraction!)



given point cloud clustered around origin: $\{\mathbb{x}_i\}_{i=1}^{N}$, where $\mathbb{x}_i = \begin{pmatrix} x_1^i \\ \vdots \\ x_K^i \end{pmatrix}$ and $N \geq K$

$$\Rightarrow \underline{\text{Covariance matrix } C} = \left(\begin{pmatrix} x_1^1 \\ \vdots \\ x_K^1 \end{pmatrix} \cdots \begin{pmatrix} x_1^N \\ \vdots \\ x_K^N \end{pmatrix} \right) \cdot \begin{pmatrix} (x_1^1 \ldots x_K^1) \\ \vdots \\ (x_1^N \ldots x_K^N) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{N} x_1^i x_1^i & \cdots & \sum_{i=1}^{N} x_1^i x_K^i \\ \vdots & & \\ \sum_{i=1}^{N} x_K^i x_1^i & \cdots & \sum_{i=1}^{N} x_K^i x_K^i \end{pmatrix}$$

(*) Define:

$$\mathbb{v}_1 = \begin{pmatrix} x_1^1 \\ \vdots \\ x_1^N \end{pmatrix}, \ldots$$

$$\mathbb{v}_K = \begin{pmatrix} x_K^1 \\ \vdots \\ x_K^N \end{pmatrix}$$

$$\overset{(*)}{=} \begin{pmatrix} \langle \mathbb{v}_1, \mathbb{v}_1 \rangle & \cdots & \langle \mathbb{v}_1, \mathbb{v}_K \rangle \\ \vdots & & \vdots \\ \langle \mathbb{v}_K, \mathbb{v}_1 \rangle & \cdots & \langle \mathbb{v}_K, \mathbb{v}_K \rangle \end{pmatrix}$$

inner product

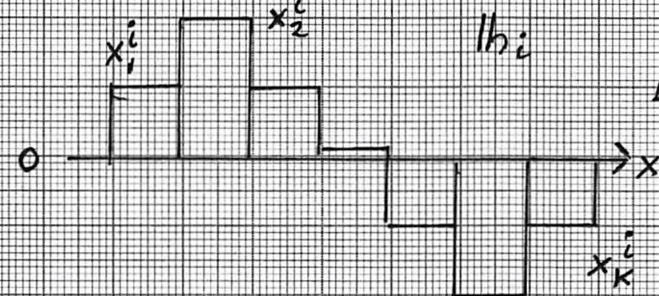
$\Rightarrow$ The K eigenvalues and -vectors define the EIGENSYSTEM of the point cloud.

"PCA defines the orthogonal eigensystem."

Stratovan

■ HISTOGRAM DISTANCES – EIGENVECTORS / EIGENFUNCTIONS OF CUMULATIVE DISTRIBUTION FUNCTIONS

2) PCA for histogram data (binned, after mean subtraction)

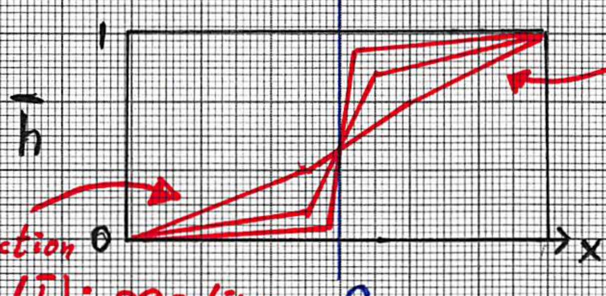


mean-subtracted histogram $\mathbb{h}_i$ viewed as point:

$$\mathbb{h}_i = \begin{pmatrix} x_1^i \\ \vdots \\ x_k^i \end{pmatrix}, \quad i = 1, \ldots, N$$

⇒ The K eigenvalues and -vectors of a set of N "K-dimensional histograms" are computed analogously to 1), see previous page.

3) PCA for cumulative distribution functions (after mean subtraction, applied to x-direction



function $x_i(\bar{h})$: ∞-dim. point

cluster of N mean-subtracted functions $x_i(\bar{h})$, $i = 1 \ldots N$, for which eigenfunctions ⊛ are computed.

⊛ Why are these eigenfunctions relevant?
Eigenfunctions associated with largest eigenvalues can be used to **GENERATE SYNTHETIC HISTOGRAMS** – e.g., needed for training, testing, ...

■ HISTOGRAM DISTANCES – PRINCIPAL COMPONENTS ANALYSIS TO CHARACTERIZE HISTOGRAMS AND DISTRIBUTION FUNCTIONS

• **IDEA:** – For point sets PCA makes it possible to define "eigendirection" based on the COVARIANCE MATRIX defined by the points.

– A "typical histogram with 64 bins" can be viewed as a point in 64-dim. space; thus, a set of such histograms can be treated by PCA as a point set.

– A cumulative histogram based on 64 original bins is also a "piecewise constant function" – that can be viewed as a point in higher-dim. space.

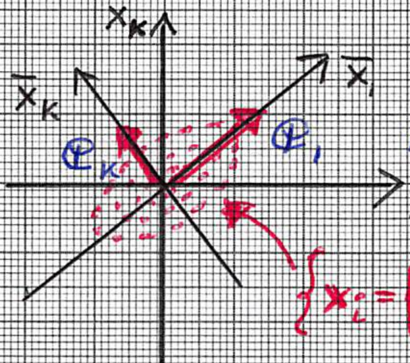
BUT: WHEN VIEWING A CUMULATIVE DISTRIBUTION (EVEN WHEN DERIVED FROM A 64-BIN HISTOGRAM) AS A "FUNCTION IN X-DIRECTION", EMBEDDING SUCH A FUNCTION AS A POINT IN HIGH-DIM. SPACE IS MORE COMPLICATED...

• **GOAL:** CHARACTERIZE HISTOGRAMS / DISTRIBUTION VIA PCA ANALYSIS APPLIED TO CUMULATIVE DISTRIBUTION FUNCTIONS "IN X-DIRECTION"

– IN SUPPORT OF GENERATING SYNTHETIC DATA WITH THE SAME DISTRIBUTION CHARACTERISTICS!

■ HISTOGRAM DISTANCES – PCA FOR CUMULATIVE DISTRIBUTION FUNCTION ANALYSIS

1) Basic principles for point data PCA:



e_1, e_2 = basis vectors proportional to eigenvectors of $\mathbf{C}$

$$\left\{ \mathbf{x}_i = \begin{pmatrix} x_1^i \\ \vdots \\ x_K^i \end{pmatrix} \right\}_{i=1}^{N}$$

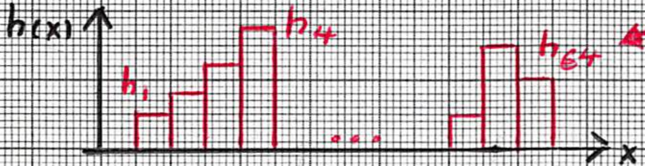
⇒ covariance matrix $\mathbf{C}$ =

$$\begin{bmatrix} x_1^1 & & x_1^N \\ & & \\ x_K^1 & & x_K^N \end{bmatrix} \begin{bmatrix} x_1^1 & & x_K^1 \\ & & \\ x_1^N & & x_K^N \end{bmatrix}, \quad N \geq K$$

($\mathbf{C}$ is a K-by-K matrix.)

• **IMPORTANT:** The "Largest eigenvalues and associated eigenvectors" of $\mathbf{C}$ define the point distribution – and can be used to generate "synthetic point sets of the same distribution.

2) Application to binned histograms:



e.g., 64 bins (K=64)

set of N such histograms: $\left\{ \mathbf{h}_i = \begin{pmatrix} h_1^i \\ \vdots \\ h_K^i \end{pmatrix} \right\}_{i=1}^{N}$

⇒ compute covariance matrix $\mathbf{C}$ for set of N K-dim. "histogram points" $\mathbf{h}_i$.

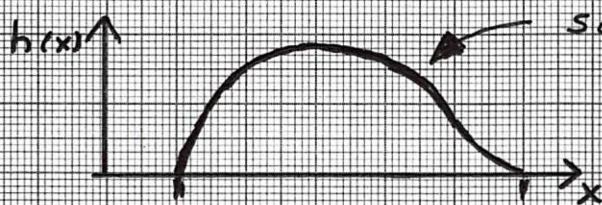
⇒ Largest eigenvalues and associated eigenvectors define a "basis" for the given set of N histograms; "small eigenvalues represent noise."

■ **HISTOGRAM DISTANCES – PCA, COVARIANCE ETC. FOR FUNCTIONS**

3) Application to general functions

(for use with cumulative distribution functions)

• Example: smooth, continuous distribution



set of N such continuous functions $h_i(x)$. Function $h_i(x)$ can be viewed as a point in ∞-dim. space:

$$h_i = \begin{pmatrix} h_1^i \\ \vdots \\ h_\infty^i \end{pmatrix}$$

continuous function

$\Rightarrow$ $K = \infty$

N = finite

$\Rightarrow$ How to compute eigenvalues and eigenvectors meaningfully for a set $\{h_i(x)\}_{i=1}^N$, using an approach based on PCA and a covariance matrix?

$\Rightarrow$ MUST TRANSITION FROM FINITE/DISCRETE CASE TO INFINITE/CONTINUOUS CASE!

4) The covariance matrices for 1), 2), 3):

C for 1):

$$C = \begin{bmatrix} \sum_{i=1}^N x_1^i x_1^i & \dots & \sum_{i=1}^N x_1^i x_K^i \\ \vdots & & \vdots \\ \sum_{i=1}^N x_K^i x_1^i & \dots & \sum_{i=1}^N x_K^i x_K^i \end{bmatrix}$$

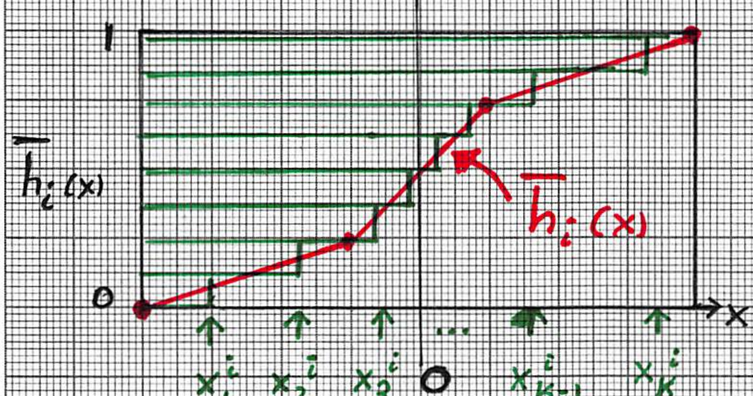
C for 2): same as 1), just replace x by h.

C for 3): $K \to \infty$ OR: $K = N$ (*)

(*) N is "Large"

HISTOGRAM DISTANCES – PCA FOR CUMULATIVE DISTRIBUTION FUNCTIONS

- Given: set of N cumulative distribution functions $\bar{h}_i(x)$, all mean-substracted
- Goal: determine largest eigenvalues and associated eigenvectors of the functions $\bar{h}_i(x)$, all given originally as piecewise linear functions; the x-direction should be used for analysis (PCA)
- Example: approximate $\bar{h}_i(x)$ – viewed as a function in x-direction – by K "uniform bars" in x-direction; e.g., $K = N$



$\Rightarrow$ the piecewise constant approximation of the cumulative function $\bar{h}_i(x)$, in x-direction, allows one to represent $\bar{h}_i(x)$ as

$$\mathbf{x}_i = \begin{pmatrix} x_1^i \\ \vdots \\ x_K^i \end{pmatrix}$$

$\Downarrow$

- PCA applied to $\{\mathbf{x}_i\}_{i=1}^{N}$ leads to covariance matrix C whose largest eigenvalues define a "basis" (the associated eigenvectors) that captures the distribution of $\{\bar{h}_i(x)\}$. $\Rightarrow$ CAN GENERATE DISTRIBUTIONS OF SAME TYPE.

~BH