BH

STRATOVAN

5/16/2016

POINT SETS – DISTRIBUTIONS –
– DISTANCE METRIC – PCA etc.

[WHY? Histograms/distributions viewed as points in high-dim. space]

- "Affine Invariant Norm" for point sets:

→ Given: n points $p_i = (x_1^i, x_2^i, \dots x_K^i)^T$, $i = 1 \dots n$, in K-dim. space

→ Norm '$\|\cdot\|$' defines square of length of vector $v = (x_1, \dots, x_K)^T$:

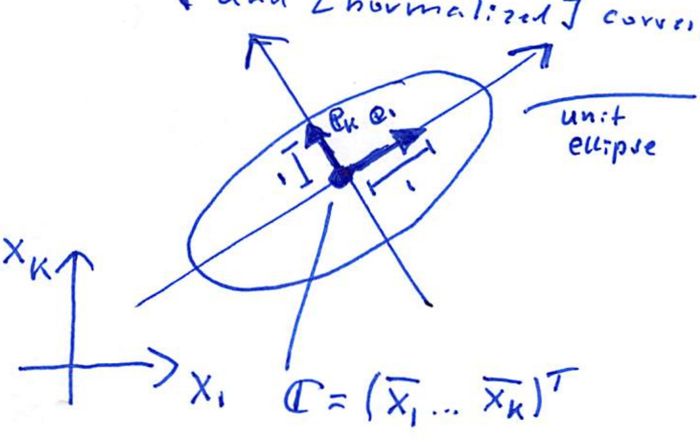
$$\|v\|^2 = v^T Q v \quad , \quad \text{where}$$

$$Q = n\,(D^T D)^{-1} \quad , \quad \text{where}$$

$$D = \begin{pmatrix} x_1^1 - \bar{x}_1 & \dots & x_K^1 - \bar{x}_K \\ \vdots & & \vdots \\ x_1^n - \bar{x}_1 & \dots & x_K^n - \bar{x}_K \end{pmatrix} , \quad \text{where}$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_j^i \quad , \quad j = 1 \dots K$$

→ Q has REAL eigenvalues $\lambda_1, \dots, \lambda_K$ (and [normalized] corresponding eigenvectors $e_1, \dots, e_K$)



→ Condition for points p on UNIT ellipsoid:

$$\boxed{p^T Q p = 1}$$

$$\boxed{l_1^2 = 1/\lambda_1, \dots, l_K^2 = 1/\lambda_K}$$

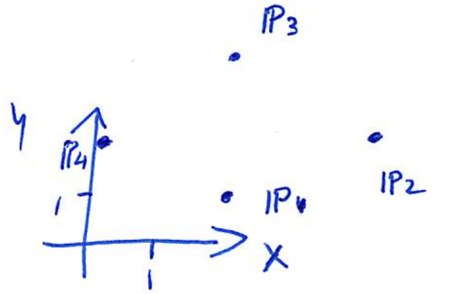
are the squared lengths of ellipsoid's principal axes.

BH

STRATOVAN

5/16/2016

• Simple 2D example

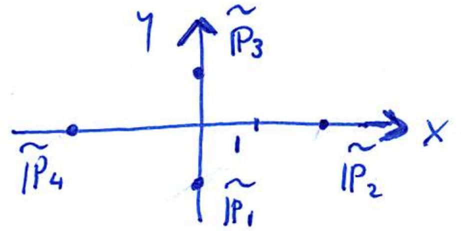


$n = 4$:

$$\mathbb{P}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbb{P}_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \mathbb{P}_3 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \mathbb{P}_4 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

(i) $$\underline{\mathbb{C}} = \frac{1}{4} \sum_{i=1}^{4} \mathbb{P}_i = \frac{1}{4}\begin{pmatrix} 8 \\ 8 \end{pmatrix} = \underline{\begin{pmatrix} 2 \\ 2 \end{pmatrix}}$$

$\Downarrow$ subtract mean $\mathbb{C}$



$$\boxed{\tilde{\mathbb{P}}_i = \mathbb{P}_i - \mathbb{C}}$$

(ii) Perform mean-subtraction and compute D ($\mathbb{P}_i \mapsto \tilde{\mathbb{P}}_i$):

$$\underline{D} = \begin{pmatrix} [& \tilde{\mathbb{P}}_1 &] \\ [& \tilde{\mathbb{P}}_2 &] \\ [& \tilde{\mathbb{P}}_3 &] \\ [& \tilde{\mathbb{P}}_4 &] \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 2 & 0 \\ 0 & 1 \\ -2 & 0 \end{pmatrix}$$

(iii) Compute Q:

$$D^T D = \begin{pmatrix} 0 & 2 & 0 & -2 \\ -1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 2 & 0 \\ 0 & 1 \\ -2 & 0 \end{pmatrix} = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\Rightarrow (D^T D)^{-1} = \begin{pmatrix} 1/8 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$\Rightarrow \underline{Q} = n \cdot (D^T D)^{-1} = 4 \cdot \begin{pmatrix} 1/8 & 0 \\ 0 & 1/2 \end{pmatrix} = \underline{\begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}}$$

(iv) Eigenvalues & eigenvectors of Q:

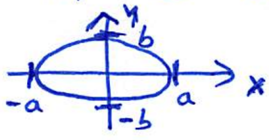
$$\begin{vmatrix} \frac{1}{2} - \lambda & 0 \\ 0 & 2 - \lambda \end{vmatrix} = 0 \Rightarrow \underline{\lambda_1 = \frac{1}{2}}, \quad \underline{\lambda_2 = 2}$$

$$\begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} e_x \\ e_y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} e_x \\ e_y \end{pmatrix} \Rightarrow \underline{e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}}; \quad \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} e_x \\ e_y \end{pmatrix} = 2\begin{pmatrix} e_x \\ e_y \end{pmatrix} \Rightarrow e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

(v) Length of ellipse's principal axes:

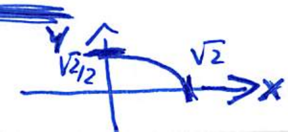
$$l_1^2 = 1/\lambda_1 = 2 \Rightarrow \underline{l_1 = \sqrt{2}}; \quad l_2^2 = 1/\lambda_2 = \frac{1}{2} \Rightarrow \underline{l_2 = \sqrt{2}/2}$$

(vi) Ellipse:



$$\boxed{\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1}$$

Here: $a = \sqrt{2}$, $b = \sqrt{2}/2$:



BH

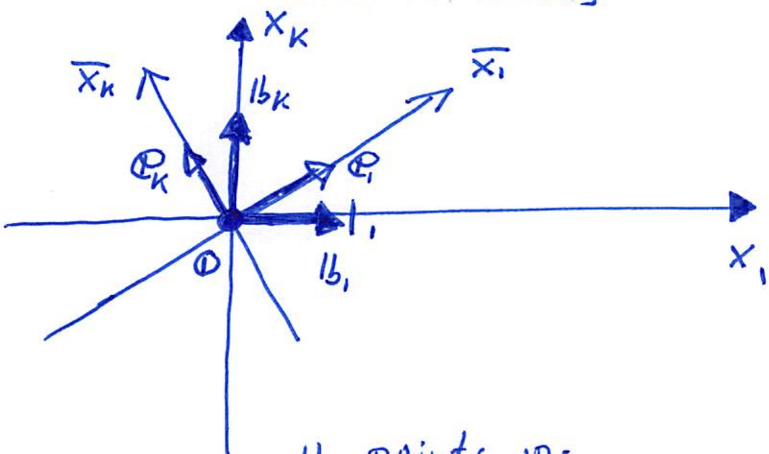
Stratovan

5/23/2016

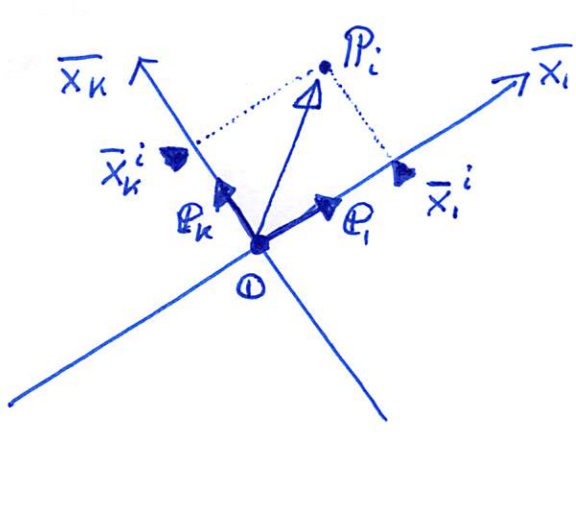
USING THE PCA-based EIGENDIRECTIONS OF A POINT SET TO DEFINE "TIGHT" OBJECT/POINT SET-ALIGNED BOUNDING BOXES (PSABBs)

- ILLUSTRATION

[MEAN SUBTRACTION HAS ALREADY BEEN DONE...]



$\Downarrow$ points $\mathbb{P}_i$ must be expressed w.r.t. eigensystem; then the MINIMAL PSABB is computed.



1) Original system: $\{\mathbb{O}, \mathbb{b}_1, \ldots, \mathbb{b}_K\}$

$\rightarrow$ coordinates $X_1, \ldots, X_K$

2) "Eigen system": $\{\mathbb{O}, \mathbb{e}_1, \ldots, \mathbb{e}_K\}$

$\rightarrow$ coordinates $\overline{X}_1, \ldots, \overline{X}_K$

3) Basis vectors $\mathbb{b}_i$ & $\mathbb{e}_i$ are all normalized.

---

4) Coordinates of $\mathbb{P}_i$ w.r.t. $\mathbb{e}_1, \ldots, \mathbb{e}_K$:

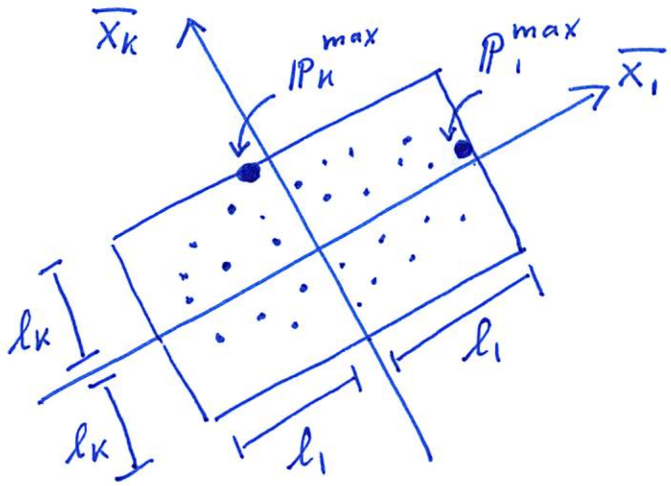
$$\overline{X}_1^i = \mathbb{P}_i \bullet \mathbb{e}_1$$
$$\vdots$$
$$\overline{X}_K^i = \mathbb{P}_i \bullet \mathbb{e}_K$$

$$\Rightarrow \overline{X}_j^i = \mathbb{P}_i \bullet \mathbb{e}_j, \quad j = 1 \ldots K \quad !$$

BH
STRATOVAN

5/23/2016

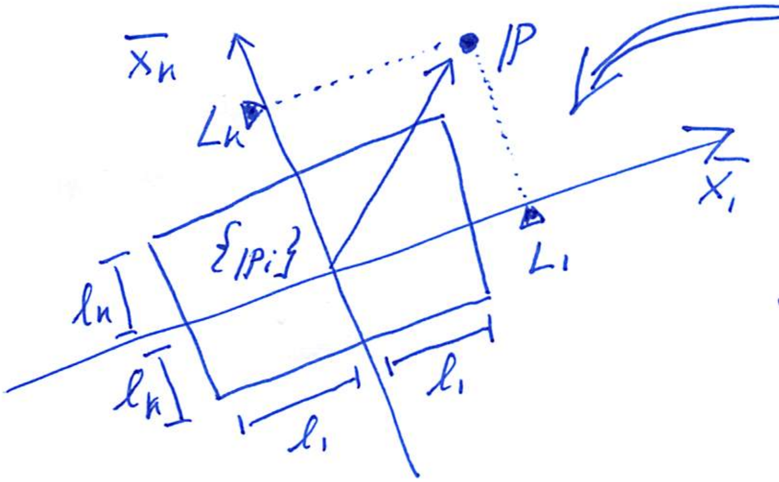
5) Establish minimal PSABB for set of $\{p_i\}$:



→ Determine the points $p_1^{max}, \ldots, p_k^{max}$ with maximal value for coordinate $\overline{x}_j$: (absolute value!)

→ minimal PSABB's edge lengths can be computed; compute values for $l_1, \ldots, l_k$.

6) Utilization: Can quickly determine whether a "new" point p (expressed relative to $e_1, \ldots, e_k$ - after mean subtraction) "is inside the point set $\{p_i\}$ or not:



- here: $|L_1| > l_1$
  $\Rightarrow$ "$p \notin \{p_i\}$"

- Can quickly "classify" point p as being of type $\{p_i\}$ - or not. ■

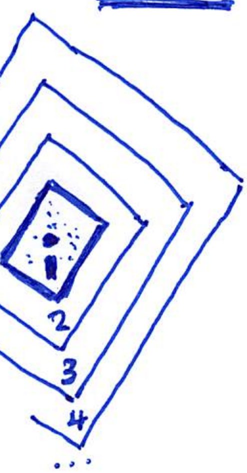
BH

BH
STRATOVAN

5/23/2016

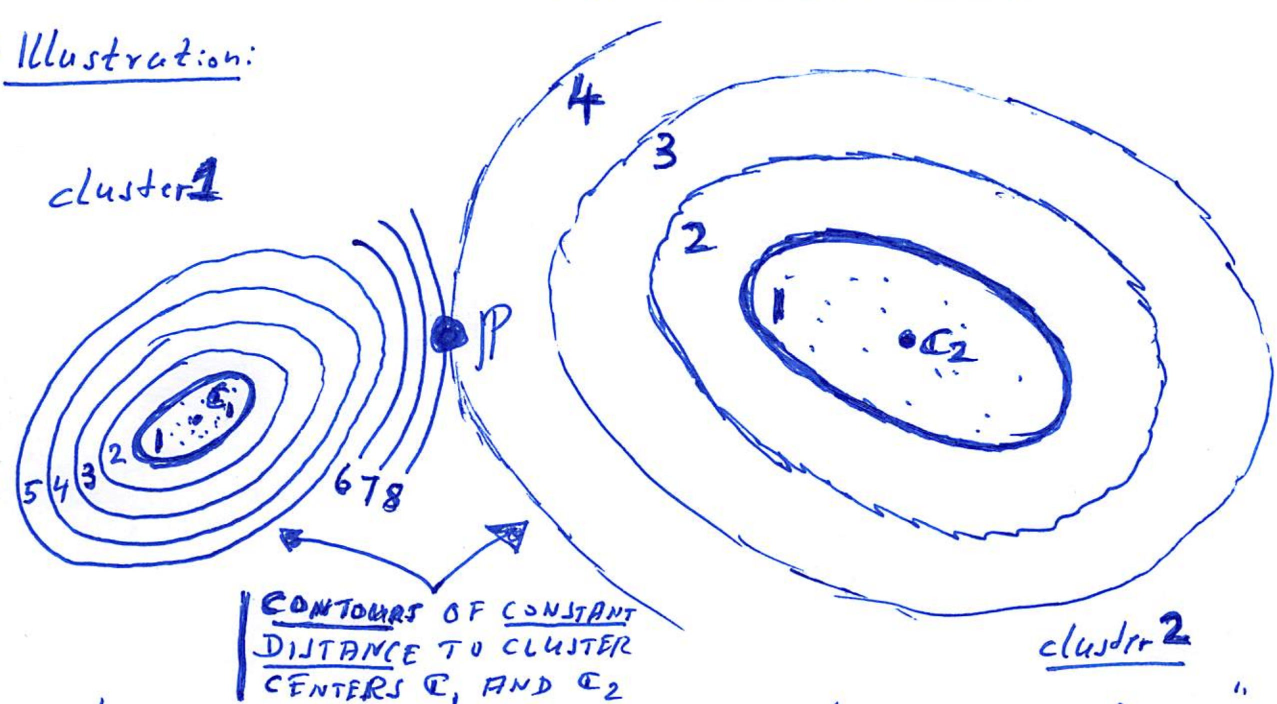
■ CAN USE POINT DENSITY & PCA-based METRIC OF DEFINED BY THE DISTINCT CLUSTERS TO ALSO DETERMINE TO WHAT POINT CLUSTER A NEW POINT $\mathbb{p}$ IS CLOSER TO:

- Idea: The eigenvalues & eigenvectors associated with a cluster define cluster-specific off-sets (or contours of 'distance from cluster center') – that can be used to define a point $\mathbb{p}$'s distance to various clusters / cluster centers:
  [SEE REF: Louis Feng, Ingrid Hotz, Bernd Hamann, Ken Joy – "Anisotropic Noise Samples"]



USE OFF-SETS OF PSABBs: EFFICIENT APPROXIMATION:

- Illustration:



here: "$\mathbb{p}$ is 8 units away from $\mathbb{c}_1$, 4 units away from $\mathbb{c}_2$";
$\Rightarrow$ $\mathbb{p}$ is closer to cluster 2 !!!

■ BH