Stratovan

DIMENSION REDUCTION – Cont'd.

• Method:... → The D eigenvalues of the covariance matrix $\mathbb{C}$ are ordered based on decreasing value, i.e.,

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_D \geq 0 \quad .$$

The associated eigenvectors $\mathbb{e}_1, \ldots, \mathbb{e}_D$, when normalized, define the desired ordered optimal orthonormal basis $\{\overline{\mathbb{b}}_i\}_{i=1}^{D}$.

→ The original coordinate system is defined by the original basis $\{\mathbb{b}_i\}_{i=1}^{D}$. The new, optimal coordinate system is defined by the orthonormal basis $\{\overline{\mathbb{b}}_i\}_{i=1}^{D}$. Thus, the system $\{\overline{\mathbb{b}}_i\}_{i=1}^{D}$ is obtained from $\{\mathbb{b}_i\}_{i=1}^{D}$ by applying the ROTATION MATRIX $R_{ot} = (\overline{\mathbb{b}}_1, \ldots, \overline{\mathbb{b}}_D)$ to $\{\mathbb{b}_i\}$.

Consequently, a point with a representation $\mathbb{x}$ relative to the original system has this "optimal" representation relative to $\{\overline{\mathbb{b}}_i\}_{i=1}^{D}$:

$$\overline{\mathbb{x}} = R_{ot}^{-1}\,\mathbb{x} = R_{ot}^{T}\,\mathbb{x} = \begin{pmatrix} \vdash & \overline{\mathbb{b}}_1 & \dashv \\ & \vdots & \\ \vdash & \overline{\mathbb{b}}_D & \dashv \end{pmatrix} \mathbb{x} \; .$$

($R_{ot}^{-1} = R_{ot}^{T}$ since R_{ot} is an orthonormal rotation matrix.)

• Ex.:

Three points in 3D space (no mean-subtraction) defining 2D subspace

→ given points/positional vectors: $\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}$

→ "covariance matrix" $\mathbb{C} = \begin{pmatrix}0&1&0\\0&0&1\\0&1&1\end{pmatrix}\begin{pmatrix}0&0&0\\1&0&1\\0&1&1\end{pmatrix} = \begin{pmatrix}1&0&1\\0&1&1\\1&1&2\end{pmatrix}$
(no mean-subtracton)

→ characteristic polynomial $pol(\lambda)$ =

$$\begin{vmatrix} 1-\lambda & 0 & 1 \\ 0 & 1-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix} = (1-\lambda)\big((1-\lambda)(2-\lambda)-2\big)$$

→ eigenvalues: $\lambda_1 = 3$, $\lambda_2 = 1$, $\lambda_3 = 0$

Stratovan

DIMENSION REDUCTION – Cont'd.

• **Ex.:** ... with proper mean-subtraction...

→ mean $= \begin{pmatrix} 1/3 \\ 1/3 \\ 2/3 \end{pmatrix}$ ⇒ points/vectors after mean-subtraction: $\begin{pmatrix} -1/3 \\ -1/3 \\ -2/3 \end{pmatrix}, \begin{pmatrix} 2/3 \\ -1/3 \\ 1/3 \end{pmatrix}, \begin{pmatrix} -1/3 \\ 2/3 \\ 1/3 \end{pmatrix}$

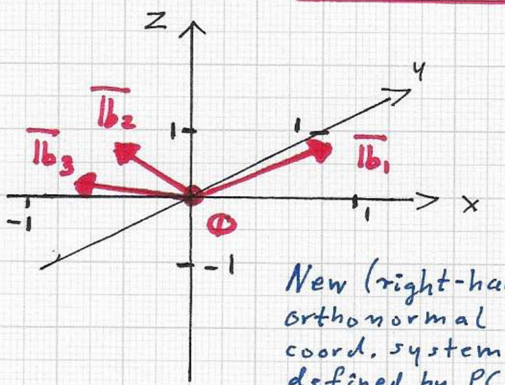
→ proper covariance matrix $\mathbf{C} = \frac{1}{9}\begin{pmatrix} 6 & -3 & 3 \\ -3 & 6 & 3 \\ 3 & 3 & 6 \end{pmatrix}$

→ eigenvalues (wolframalpha.com): $\lambda_1 = 1,\ \lambda_2 = 1,\ \lambda_3 = 0$

(example showing necessity of mean-subtraction!)

→ normalized eigenvectors:

$$\overline{\mathbf{b}}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} / \sqrt{2}, \quad \overline{\mathbf{b}}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} / \sqrt{2}, \quad \overline{\mathbf{b}}_3 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} / \sqrt{3}$$



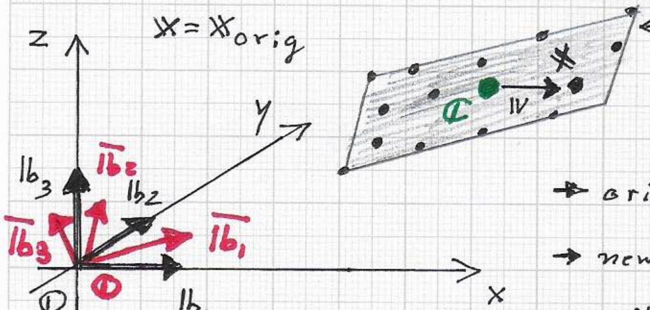
New (right-handed) orthonormal coord. system defined by PCA

⇒ $\overline{\mathbf{b}}_1 \perp \overline{\mathbf{b}}_2,\ \overline{\mathbf{b}}_1 \perp \overline{\mathbf{b}}_3,\ \overline{\mathbf{b}}_2 \perp \overline{\mathbf{b}}_3$

⇒ $\overline{\mathbf{b}}_1$ and $\overline{\mathbf{b}}_2$ span the relevant 2D space, i.e., the plane passing through the origin $\mathbf{0}$.

⇒ SINCE $\lambda_3 = 0$, THE THIRD DIMENSION DEFINED BY $\overline{\mathbf{b}}_3$ IS IRRELEVANT.

• **Note:** Geometrical meaning of this "optimal" coordinate transformation



2D manifold (plane) containing all points $\mathbf{x}$ ('•')

→ orig. coord. system: $\{\mathbf{0}, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$

→ new optimal coord. system: $\{\mathbf{0}, \overline{\mathbf{b}}_1, \overline{\mathbf{b}}_2, \overline{\mathbf{b}}_3\}$

→ all original points $\mathbf{x}$ undergo mean-subtraction, i.e., $\mathbf{x} \mapsto \mathbf{x} - \mathbb{C} = \mathbf{v}$ ($\mathbb{C}$ being mean)

→ PCA applied to all vectors $\mathbf{v}$

$\mathbf{x}_{orig} = x\,\mathbf{b}_1 + y\,\mathbf{b}_2 + z\,\mathbf{b}_3$

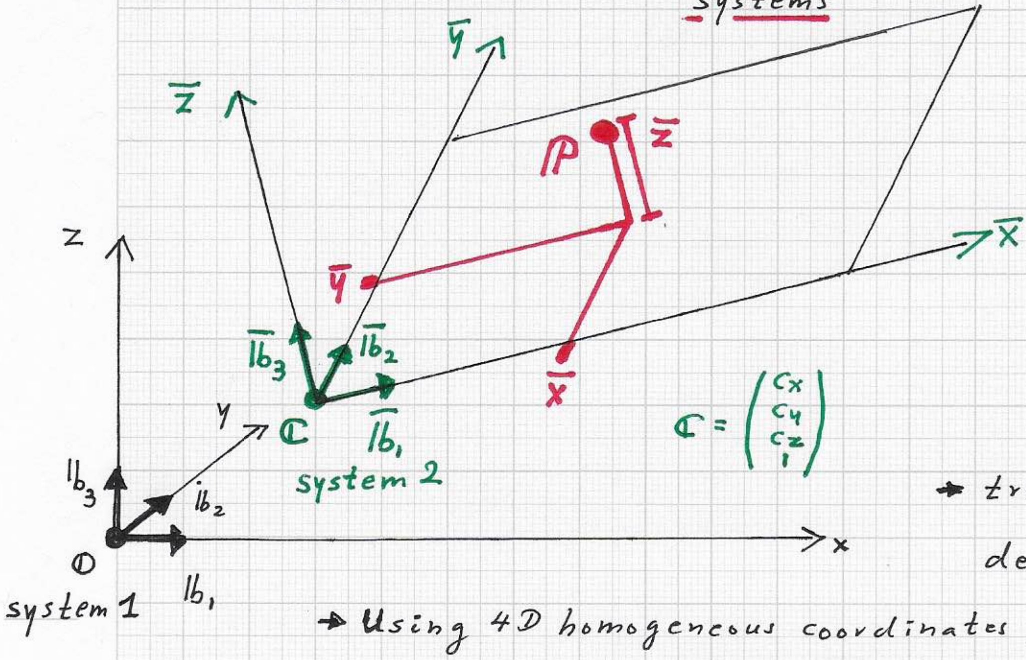
$\overline{\mathbf{x}} = \mathbf{x}_{opt} = \mathbb{C} + \overline{x}\,\overline{\mathbf{b}}_1 + \overline{y}\,\overline{\mathbf{b}}_2$

($\overline{z} = 0$ here, since $\lambda_3 = 0$)

~ BH

DIMENSION REDUCTION – Cont'd.

• **Geometry:** 2 orthonormal coordinate systems



→ system 1: $\{\mathbb{O}, \mathbb{b}_1, \mathbb{b}_2, \mathbb{b}_3\}$

→ system 2: $\{\mathbb{C}, \bar{\mathbb{b}}_1, \bar{\mathbb{b}}_2, \bar{\mathbb{b}}_3\}$

→ rotation matrix R defines mapping $\mathbb{b}_i \mapsto \bar{\mathbb{b}}_i$:

$$R = \begin{pmatrix} | & | & | \\ \bar{\mathbb{b}}_1, & \bar{\mathbb{b}}_2, & \bar{\mathbb{b}}_3 \\ | & | & | \end{pmatrix}$$

→ translation vector $\mathbb{C} - \mathbb{O} = \vec{\mathbb{C}}$ defines mapping $\mathbb{O} \mapsto \mathbb{C}$

→ Using 4D homogeneous coordinates, point $\mathbb{P}$ has the (column vector) representation $(x, y, z, 1)^T$ relative to system 1: $\mathbb{P}_{sys1} = (x, y, z, 1)^T$.

→ Point $\mathbb{P}$ has the following representation relative to system 2 (concatenation of inverses of translation and rotation):

$$\mathbb{P}_{sys2} = \begin{pmatrix} — & \bar{\mathbb{b}}_1 & — & 0 \\ — & \bar{\mathbb{b}}_2 & — & 0 \\ — & \bar{\mathbb{b}}_3 & — & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbb{P}_{sys1}$$

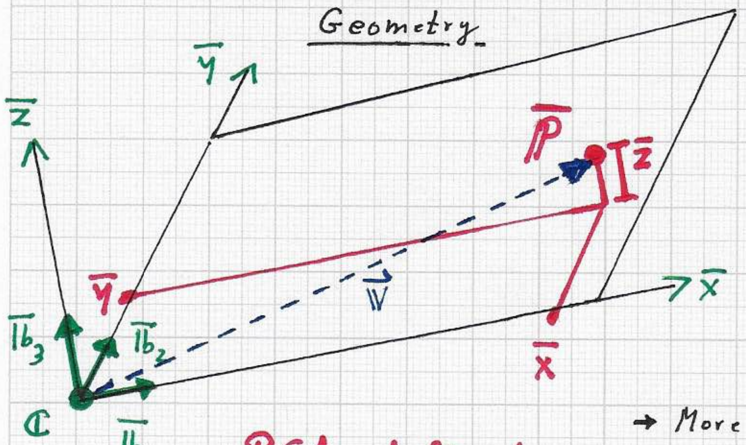
$$= (\bar{x}, \bar{y}, \bar{z}, 1)^T$$

• **Context:** We are given a set of points $\mathbb{P}$ to which PCA is applied (after subtraction of the mean $\mathbb{C}$). PCA generates three eigenvalues, $\lambda_1, \lambda_2, \lambda_3$, with ordered ("prioritized") orthonormal basis vectors $\bar{\mathbb{b}}_1, \bar{\mathbb{b}}_2, \bar{\mathbb{b}}_3$. Relative to system 2 ($\mathbb{C}$ being the new origin, $\bar{\mathbb{b}}_1, \bar{\mathbb{b}}_2, \bar{\mathbb{b}}_3$ being the new unit basis vectors) $\bar{\mathbb{P}} = \bar{x}\,\bar{\mathbb{b}}_1 + \bar{y}\,\bar{\mathbb{b}}_2 + \bar{z}\,\bar{\mathbb{b}}_3$.

DIMENSION REDUCTION – Cont'd.

- Dimension reduction:

Geometry



PCA-defined representation of $\overline{p}$

→ Relative to the system $\{\mathbb{C}, \overline{b}_1, \overline{b}_2, \overline{b}_3\}$ the coordinate representation of '•' is given as $\overline{p} = \overline{x}\,\overline{b}_1 + \overline{y}\,\overline{b}_2 + \overline{z}\,\overline{b}_3$, considering the point $\mathbb{C}$ as new local origin.

→ More precisely, the values of $\overline{x}, \overline{y}, \overline{z}$ are the lengths of the projections of the positional vector $\vec{v}$ onto $\overline{b}_1, \overline{b}_2, \overline{b}_3$.

→ By construction, the representations of all points $\overline{p}$ in the data set to be analyzed are optimal (in a statistical, averaged sense). The basis vectors $\overline{b}_i$ are ordered based on their statistical significance – i.e. $\overline{x}$ is (statistically) the most important coordinate and $\overline{z}$ is the least important coordinate.

→ **Dimension and data reduction is achieved by considering only some of the most important coordinates of $\overline{p}$ (for subsequent processing, analysis and, ultimately, classification).**

→ An (approximation) error measure is needed to assess the quality of only a partial expansion of $\overline{p}$ after truncating terms.

- Example: $L = \|v\| = \sqrt{\overline{x}^2+\overline{y}^2+\overline{z}^2}$ /* Length of positional vector v */

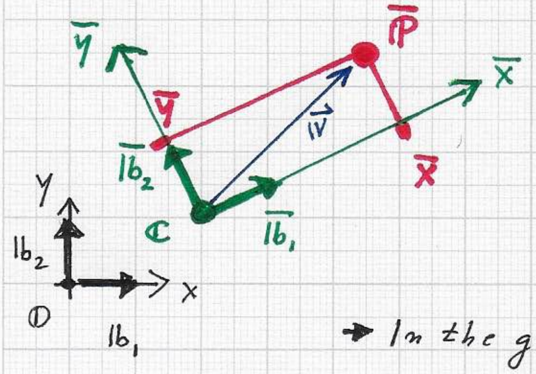
$E_1 = | L - \sqrt{\overline{x}^2} |$, /* errors E_i for expansions */

$E_2 = | L - \sqrt{\overline{x}^2+\overline{y}^2} |$, /* using the first i terms */

$E_3 = | L - \sqrt{\overline{x}^2+\overline{y}^2+\overline{z}^2} | = 0$

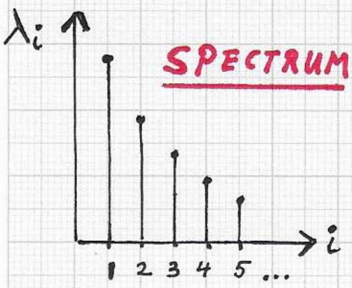
DIMENSION REDUCTION – Cont'd.

• PCA-based expansion, approximation error, truncation criteria:



→ $\overline{\mathbb{P}}$ understood (equivalently) as positional vector $\vec{\mathbb{V}}$, relative to system $\{\mathbb{C}, \overline{\mathbb{b}}_1, \overline{\mathbb{b}}_2\}$ has the expansion $\overline{\mathbb{P}} = \overline{x}\,\overline{\mathbb{b}}_1 + \overline{y}\,\overline{\mathbb{b}}_2$.

→ In the general D-dimensional case $\overline{\mathbb{P}} = \sum_{i=1}^{D} \overline{x}_i \overline{\mathbb{b}}_i$.



Since $\lambda_1 \geq \ldots \geq \lambda_D \geq 0$ and basis vectors $\overline{\mathbb{b}}_i$ inherit the index of the ordered eigenvalues λ_i, $\overline{x}_1$ is the most important and $\overline{x}_D$ the least important coordinate of $\overline{\mathbb{P}}$. Question: How many of $\overline{\mathbb{P}}$'s coordinates suffice?

→ Eigenvalues $\lambda_1, \ldots, \lambda_D$ result from minimizing a mean-squared-error-based cost function, and $\overline{\mathbb{P}}$'s approximation using the first terms, i.e., $\overline{\mathbb{P}}_R = \sum_{i=1}^{R} \overline{x}_i \overline{\mathbb{b}}_i$, has the error $E_R = \sum_{i=R+1}^{D} \lambda_i$ as associated R-term mean-squared error.

→ Statistics defines the energy Λ, based on the eigenvalues: $\Lambda = \sum_{i=1}^{D} \lambda_i$. Thus, an R-term approximation captures the energy $\Lambda_R = \sum_{i=1}^{R} \lambda_i$. The normalized R-term approximation energy is

$$\Lambda_R^{norm} = \frac{\Lambda_R}{\Lambda} = \frac{\sum_{i=1}^{R} \lambda_i}{\sum_{i=1}^{D} \lambda_i}.$$

The value of Λ_R^{norm} can be used to determine whether R terms suffice.

BH