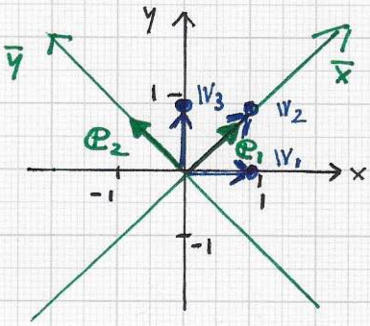
■ DIMENSION REDUCTION - Cont'd.

● PCA and its usefulness to identify important dimensions

• Ex.: PCA producing different eigenvalues, all $\neq 0$



- given positional vectors: $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$
- covariance matrix $\mathbb{C}$ =

$$= \begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 1 & 1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}$$

- characteristic polynomial

$$pol(\lambda) = \begin{vmatrix}2-\lambda & 1\\ 1 & 2-\lambda\end{vmatrix} = (2-\lambda)^2 - 1$$

- eigenvalues: $(2-\lambda)^2 = 1 \Rightarrow 2-\lambda = +/- 1$

$$\lambda_1 = 3, \quad \lambda_2 = 1$$

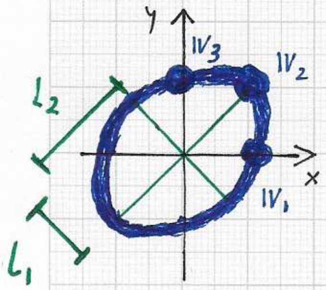
- eigenvectors: $\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \lambda\begin{pmatrix}x\\y\end{pmatrix}$

$$\Rightarrow \mathbb{e}_1 = \begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2}, \quad \mathbb{e}_2 = \begin{pmatrix}-1\\1\end{pmatrix}/\sqrt{2}$$

STATISTICS ↑

GEOMETRY ↓

•Note: Can use QUADRATIC FORM to perform analysis



- construct implicitly defined ellipse as

$$f(x,y) = c_{20}x^2 + c_{11}xy + c_{02}y^2 = 1$$

- here: 3 equations, $f(x_i, y_i) = \ldots = 1$, $i = 1\ldots3$
- solution of $c_{20}x_i^2 + c_{11}x_iy_i + c_{02}y_i^2 = 1$, $i = 1\ldots3$, is coefficient vector $c_{20} = 1$, $c_{11} = -1$, $c_{02} = 1$
- ellipse: $1x^2 - 1xy + 1y^2 = 1$

$$\text{or}: \quad (x, y)\begin{pmatrix}1 & -1/2\\ -1/2 & 1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = 1$$

$$\mathbb{x}^T \quad \mathbb{Q} \quad \mathbb{x} = 1$$

$L_1^2 = 1/\lambda_1 = 2/3$

$\Rightarrow L_1 = \sqrt{6}/3$

$L_2^2 = 1/\lambda_2 = 2$

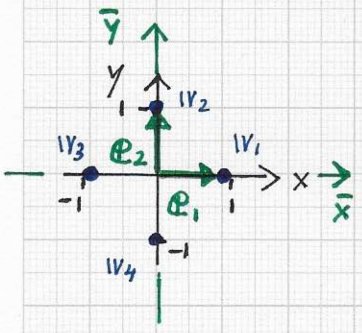
$\Rightarrow L_2 = \sqrt{2}$

- eigenvalues and eigenvectors of $\mathbb{Q}$:

$$\lambda_1 = 3/2, \ \lambda_2 = 1/2; \ \mathbb{e}_1 = \begin{pmatrix}-1\\1\end{pmatrix}/\sqrt{2}, \ \mathbb{e}_2 = \begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2}$$

■ DIMENSION REDUCTION – Cont'd.

• **Ex.:** PCA producing eigenvalues with "multiplicities > 1", all $\neq 0$

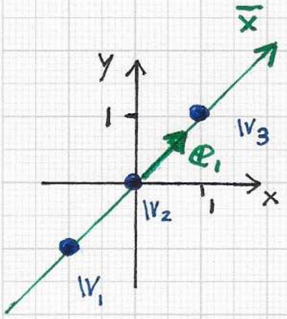


- given positional vectors: $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}-1\\0\end{pmatrix}, \begin{pmatrix}0\\-1\end{pmatrix}$
- covariance matrix C =

$$\begin{pmatrix}1 & 0 & -1 & 0\\0 & 1 & 0 & -1\end{pmatrix}\begin{pmatrix}1 & 0\\0 & 1\\-1 & 0\\0 & -1\end{pmatrix} = \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}$$

- characteristic polynomial $pol(\lambda) = (2-\lambda)^2$
- eigenvalues: $(2-\lambda)^2 = 0 \Rightarrow \lambda_1 = \lambda_2 = 2$
- eigenvectors: pair of orthogonal vectors can be chosen, e.g., $e_1 = \begin{pmatrix}1\\0\end{pmatrix}, e_2 = \begin{pmatrix}0\\1\end{pmatrix}$

• **Ex.:** PCA producing eigenvalues being 0



- given positional vectors: $\begin{pmatrix}-1\\-1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}$
- covariance matrix C =

$$\begin{pmatrix}-1 & 0 & 1\\-1 & 0 & 1\end{pmatrix}\begin{pmatrix}-1 & -1\\0 & 0\\1 & 1\end{pmatrix} = \begin{pmatrix}2 & 2\\2 & 2\end{pmatrix}$$

- characteristic polynomial $pol(\lambda) = (2-\lambda)^2 - 4$
- eigenvalues: $(2-\lambda)^2 = 4 \Rightarrow \lambda_1 = 4,\ \lambda_2 = 0$
- eigenvectors: $e_1 = \begin{pmatrix}1\\1\end{pmatrix} / \sqrt{2}$

$\Rightarrow$ The covariance matrix C has real, non-negative eigenvalues. The large eigenvalues and their associated eigenvectors define data-inherent dimensions that represent large "information content." Transforming data to these dimensions is the goal.

DIMENSION REDUCTION – Cont'd.

• Goal: [Given is a set of feature points/positional vectors – after mean subraction applied to all points – represented in D-dimensional space. They are written as $(x_1^i, \ldots, x_D^i)^T$, $i = 1 \ldots N$.]

For a set of data $\{\mathbf{x}_i\}_{i=1}^N$, originally represented in some orthogonal coordinate system, determine AN OPTIMAL ORTHOGONAL TRANSFORMATION such that the representation of the data in the resulting orthogonal basis is optimal.

• Idea:
- → Point $\mathbf{x}$ is given relative to original basis vectors $\mathbf{b}_1, \ldots, \mathbf{b}_D$, i.e., $\mathbf{x} = x_1 \mathbf{b}_1 + \ldots + x_D \mathbf{b}_D$.
- → Point $\mathbf{x}$ is undergoing an orthogonal coordinate transformation, where the point is now represented relative to new basis vectors $\overline{\mathbf{b}}_1, \ldots, \overline{\mathbf{b}}_D$, i.e., $\overline{\mathbf{x}} = \overline{x}_1 \overline{\mathbf{b}}_1 + \ldots + \overline{x}_D \overline{\mathbf{b}}_D$.
- → **Instead of considering all D coordinates of $\overline{\mathbf{x}}$ consider only (the first) R coordinates, i.e., $\overline{\mathbf{x}}^R = \overline{x}_1 \overline{\mathbf{b}}_1 + \ldots + \overline{x}_R \overline{\mathbf{b}}_R$, $R < D$.**
- → Point $\overline{\mathbf{x}}^R$ is an approximation of point $\mathbf{x}$. The approximation error is defined as $\| \mathbf{x} - \overline{\mathbf{x}}^R \|$.
- → **OPTIMIZATION PROBLEM: Considering all N data, minimize the sum of squared errors, i.e.,**

$$\sum_{i=1}^{N} \| \mathbf{x}_i - \overline{\mathbf{x}}_i^R \|^2 \Rightarrow MIN. \quad \Rightarrow \text{FIND BASIS!}$$

DIMENSION REDUCTION - Cont'd.

- **Method:** The (unknown) optimal mutually orthogonal and normalized basis vectors $\{\overline{\mathbb{b}}_i\}_{i=1}^{D}$ define the transformed representation of a point as

$$\overline{\mathbb{x}} = \underbrace{\overline{x}_1 \overline{\mathbb{b}}_1 + \ldots + \overline{x}_R \overline{\mathbb{b}}_R}_{= \overline{\mathbb{x}}^P} + \underbrace{\overline{x}_{R+1} \overline{\mathbb{b}}_{R+1} + \ldots + \overline{x}_D \overline{\mathbb{b}}_D}_{= \text{error}} .$$

An optimal orthonormal basis $\{\overline{\mathbb{b}}_i\}_{i=1}^{D}$ minimizes the sum of all squared (truncation) errors, considering therefore the errors of all points $\overline{\mathbb{x}}_j$, $j = 1 \ldots N$. Further, the terms in the expansion of $\overline{\mathbb{x}}$ are "ordered based on importance / error contribution," with $\overline{x}_1 \overline{\mathbb{b}}_1$ being the most important and $\overline{x}_D \overline{\mathbb{b}}_D$ being the least important.

$\Rightarrow$ Algorithm:

- Find best 1D subspace S_1 (best $\overline{\mathbb{b}}_1$).
- Find the next best 1D subspace S_2 (best $\overline{\mathbb{b}}_2$), where $S_1 \perp S_2$ (orthogonality).
- Iterate: Find next best 1D subspace S_i (best $\overline{\mathbb{b}}_i$), where $W_i \perp W_j$ FOR ALL $j < i$.

Conditions for $\overline{\mathbb{b}}_1$:

1) $\overline{\mathbb{b}}_1$ maximizes the mean (average) squared projections of all points onto itself ($\overline{\mathbb{b}}_1$) (= squared inner products of all points and $\overline{\mathbb{b}}_1$), i.e.,

$$\sum_{i=1}^{N} \langle \overline{\mathbb{b}}_1, \mathbb{x}_i \rangle^2 \rightarrow MAX \qquad \text{(extremal condition)}$$

2) $\| \overline{\mathbb{b}}_1 \| = 1$, i.e., $\sqrt{\langle \overline{\mathbb{b}}_1, \overline{\mathbb{b}}_1 \rangle} = 1$ (normalization condition)

■ DIMENSION REDUCTION - Cont'd.

• Method: ... The conditions for the best basis vector $\overline{\mathbb{b}}_1$ define a CONSTRAINED OPTIMIZATION problem, where 1) defines the (error) function to be made extremal and 2) defines the constraint. The LAGRANGIAN method allows one to determine the necessary conditions to be satisfied by the unknown coordinates of $\overline{\mathbb{b}}_1$ in D-dimensional space, while ensuring that $\overline{\mathbb{b}}_1$ is of length one. ... Analogously, one determines the equations to be satisfied by the remaining unknown basis vectors $\overline{\mathbb{b}}_i$, $i=2 \ldots D$. ...

The result:

The optimal basis vectors $\overline{\mathbb{b}}_i$ are defined by the solution of the EIGENVECTOR PROBLEM

$$\mathbf{C}\,\overline{\mathbb{b}}_i = \lambda_i\,\overline{\mathbb{b}}_i \quad , \quad i=1 \ldots D.$$

$\mathbf{C}$ is the symmetric covariance matrix given by all data, i.e., $\mathbf{C} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)\begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix}$,

having real, non-negative eigenvalues.

Note: → Eigenvalues with value zero have associated eigenvectors that can be ignored.

→ Eigenvalues can have multiplicities greater than one. In this case, the needed number of orthonormal eigenvectors, not uniquely defined, can be chosen to span the respective space.