# Outline

# Finding the best axis (1)

## Definition of the problem

- The best axis $\Delta_1$ is the "closest" to the point cloud
- Inertia of $\Delta_1$ measures the distance between the data and $\Delta_1$
- $\Delta_1$ is defined by the director vector $\mathbf{u}_1$, such as $\|\mathbf{u}_1\| = 1$
- $\Delta_1^\perp$ is defined by the normal vector $\mathbf{u}_1$, such as $\|\mathbf{u}_1\| = 1$

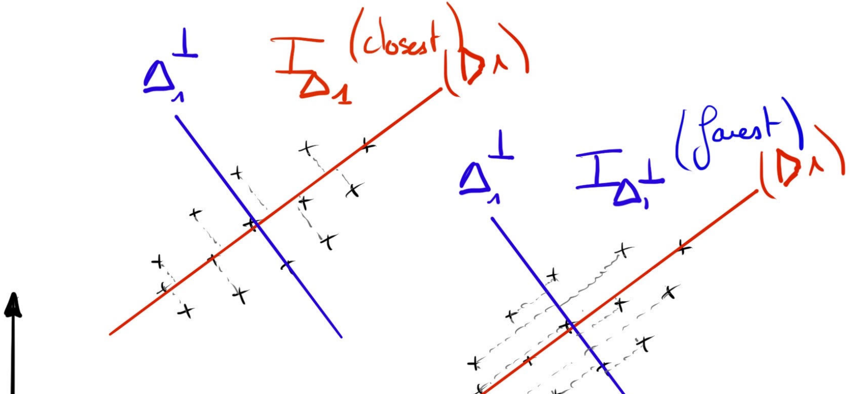
$\rightsquigarrow$ The best axis $\Delta_1$ is the one with the minimal Inertia.

# Finding the best axis (2)

## Stating the optimization problem

Since $\Delta_1 \oplus \Delta_1^\perp = \mathbb{R}^p$ and $I_T = I_{\Delta_1} + I_{\Delta_1^\perp}$, then

$$\underset{\mathbf{u}\in\mathbb{R}^p\,:\,\|\mathbf{u}\|=1}{\text{minimize}}\; I_{\Delta_1} \Leftrightarrow \underset{\mathbf{u}\in\mathbb{R}^p\,:\,\|\mathbf{u}\|=1}{\text{maximize}}\; I_{\Delta_1^\perp}$$



# Finding the best axis (3)

## Stating the problem (algebraically)

Find $\mathbf{u}_1; \|\mathbf{u}_1\| = 1$ that maximizes

$$
\begin{aligned}
I_{\Delta_1^\perp} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{dist}(\mathbf{x}_i, \Delta_1^\perp)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbf{u}_1^\top(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{u}_1 \\
&= \mathbf{u}_1^\top \left( \sum_{i=1}^{n} \frac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \right) \mathbf{u}_1 \\
&= \mathbf{u}_1^\top \hat{\boldsymbol{\Sigma}} \mathbf{u}_1
\end{aligned}
$$

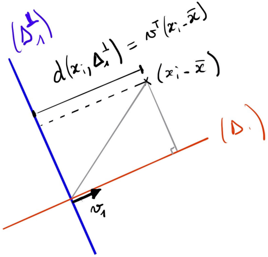



Figure 1: Geometrical insight

# Finding the best axis (4)

We solve a simple constraint maximization problem with the method of Lagrange multipliers:

By straightforward (vector) differentiation, an using that $\mathbf{u}_1^\top \mathbf{u}_1 = 1$

# Finding the following axes

Second best axis

# Interpretation in $\mathbb{R}^p$

$\mathbf{U}$ describes a new orthogonal basis and a rotation of data in this basis

⇝ PCA is an appropriate rotation on axes that maximizes the variance

$$\begin{cases} \Delta_1 & \oplus & \dots & \oplus & \Delta_p \\ \mathbf{u}_1 & \perp & \dots & \perp & \mathbf{u}_p \\ \lambda_1 & > & \dots & > & \lambda_p \\ I_{\Delta_1^\perp} & > & \dots & > & I_{\Delta_p^\perp} \end{cases}$$