

Tout savoir sur l'ACP

Partie 1 : La théorie

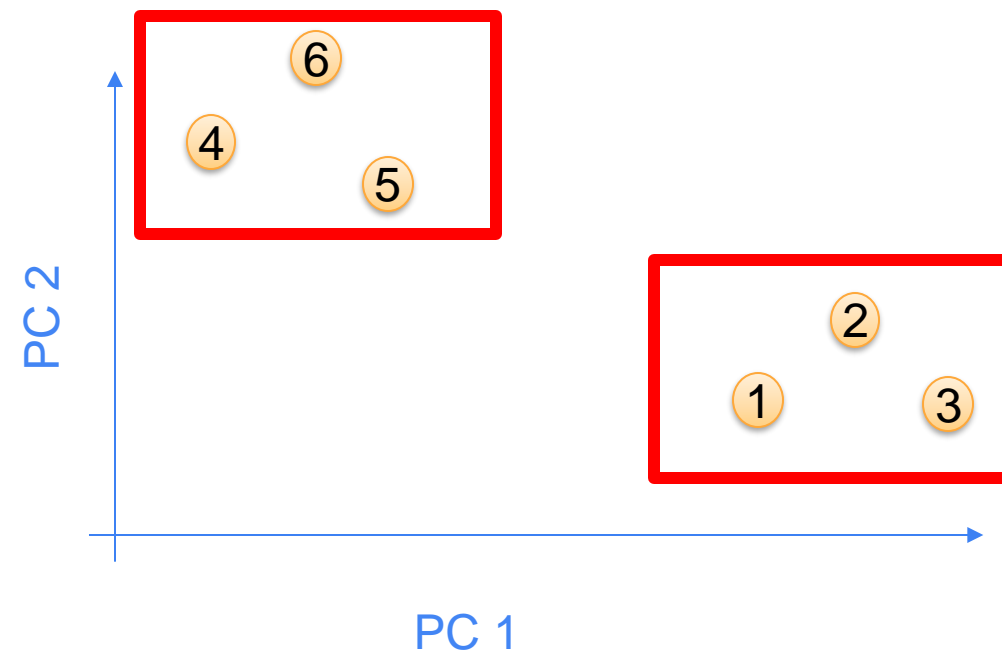


Présenté par **Morgan Gautherot**



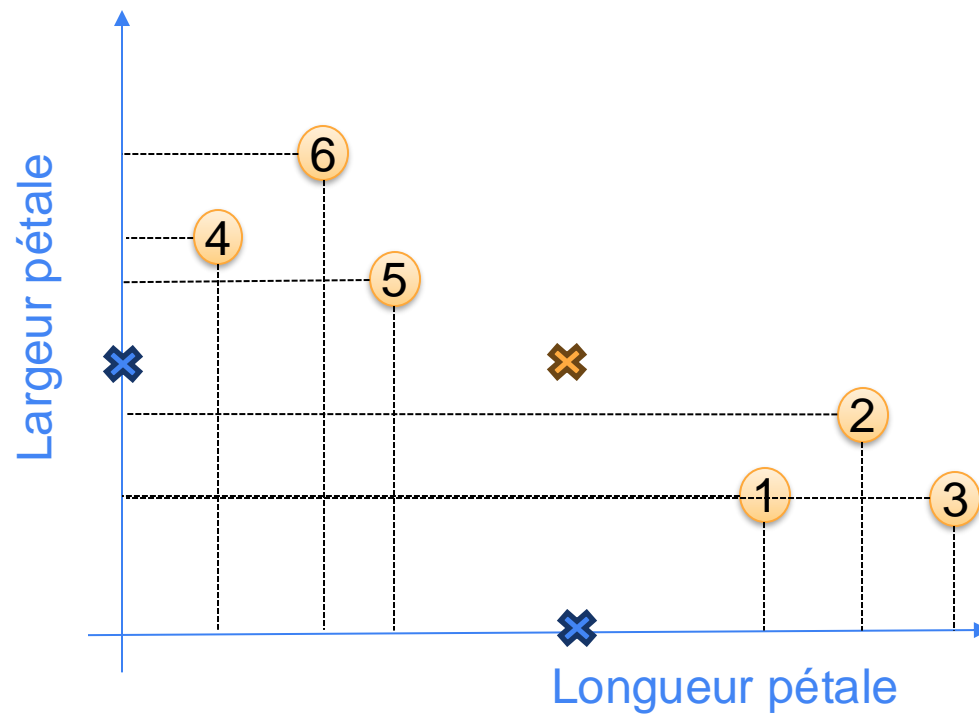
L'objectif de l'ACP

	Longueur pétale	Largeur pétale	Longueur sepal	Largeur sepal	Longueur tige	...
Fleur_1	11	2	12
Fleur_2	15	5	11
Fleur_3	16	2	14
Fleur_4	3	10	8
Fleur_5	5	9	9
Fleur_6	4	12	0
...



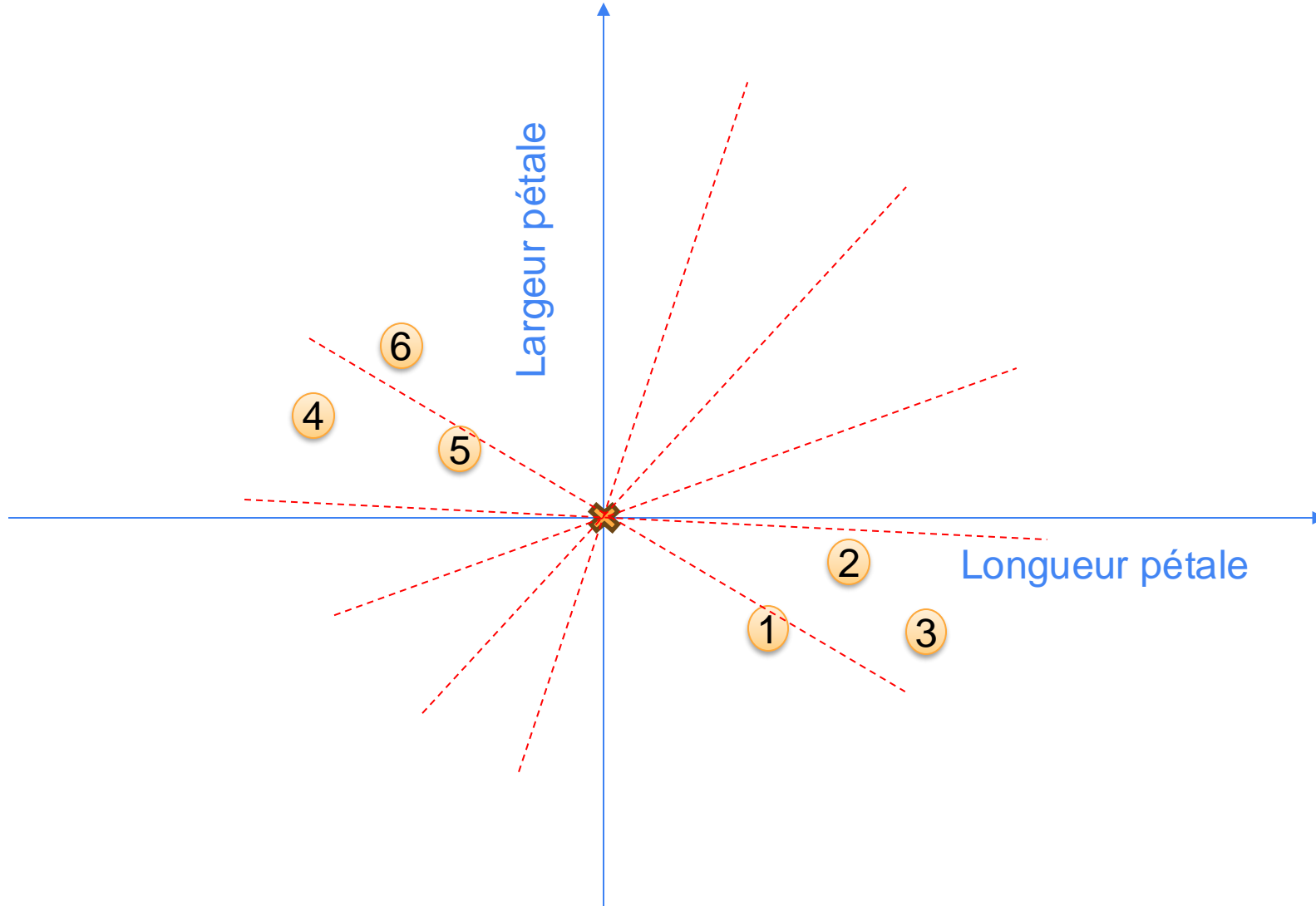


Centrer les données





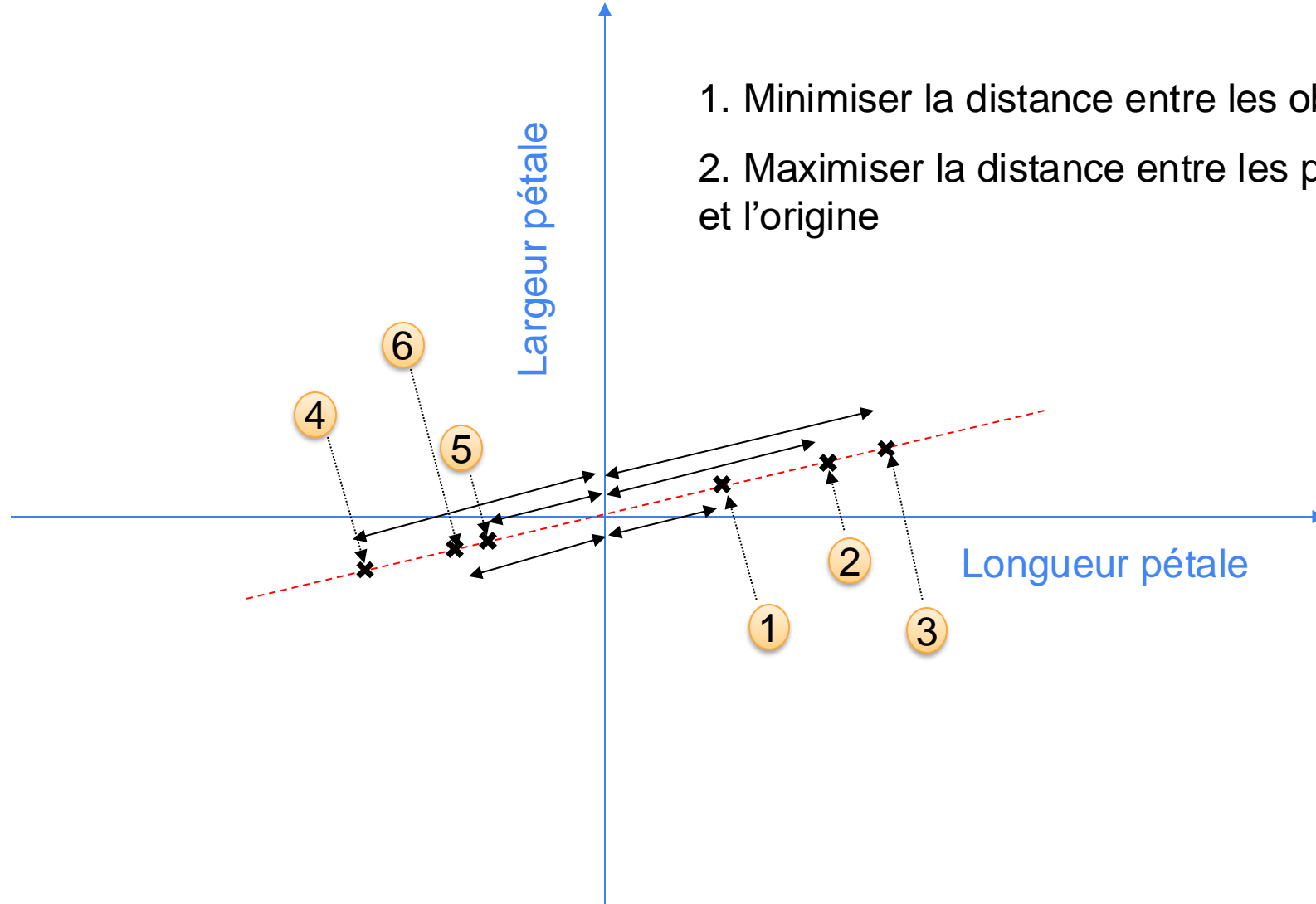
Centrer les données





Comment trouver la meilleure droite ?

1. Minimiser la distance entre les observations et la droite
2. Maximiser la distance entre les projections des observations et l'origine





Comment trouver la meilleure droite ?

Si b^2 devient plus grand c^2 devient plus petit

Si c^2 devient plus grand b^2 devient plus petit

Selon le théorème de Pythagore

$$a^2 = b^2 + c^2$$

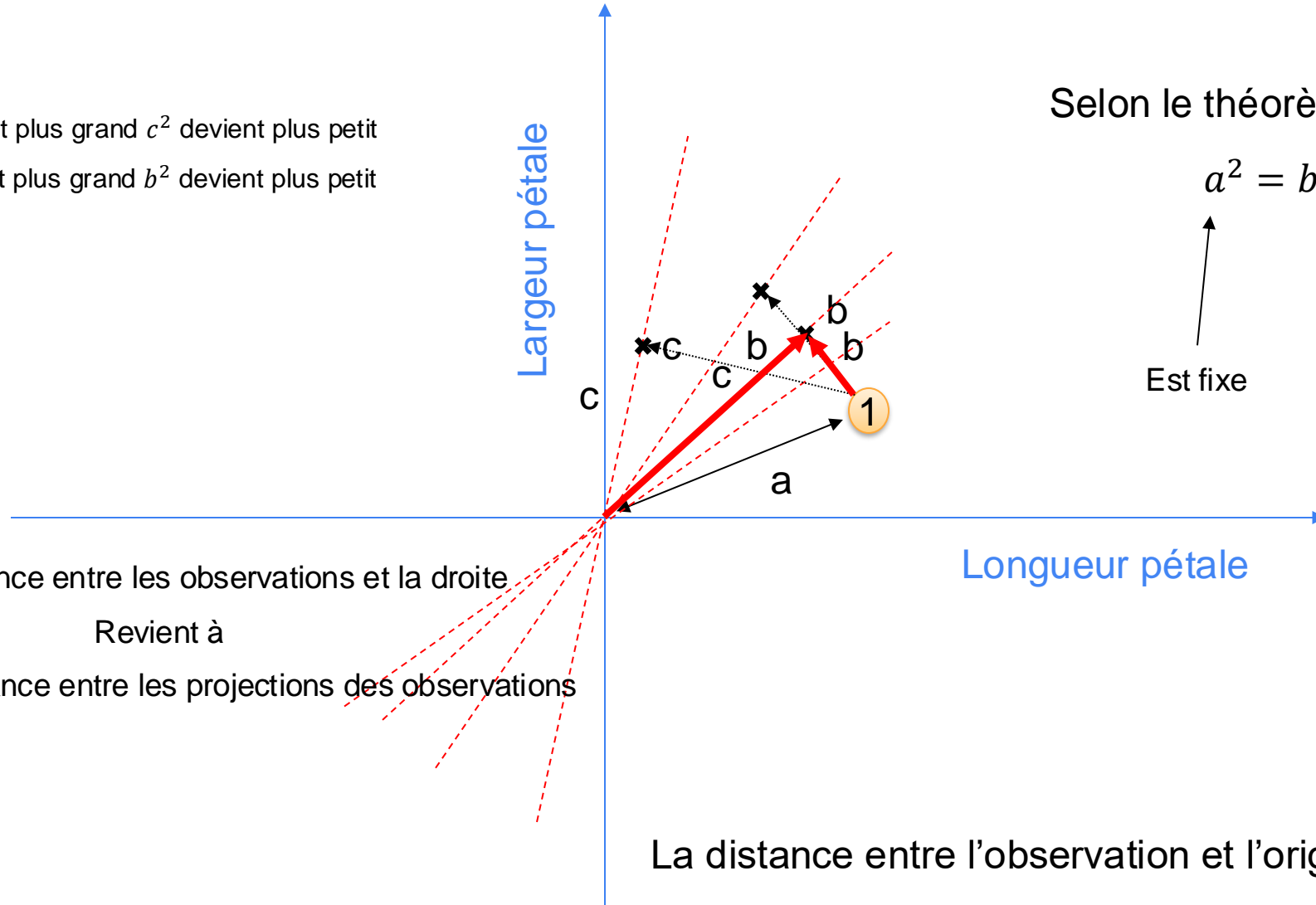
Est fixe

Minimiser la distance entre les observations et la droite

Revient à

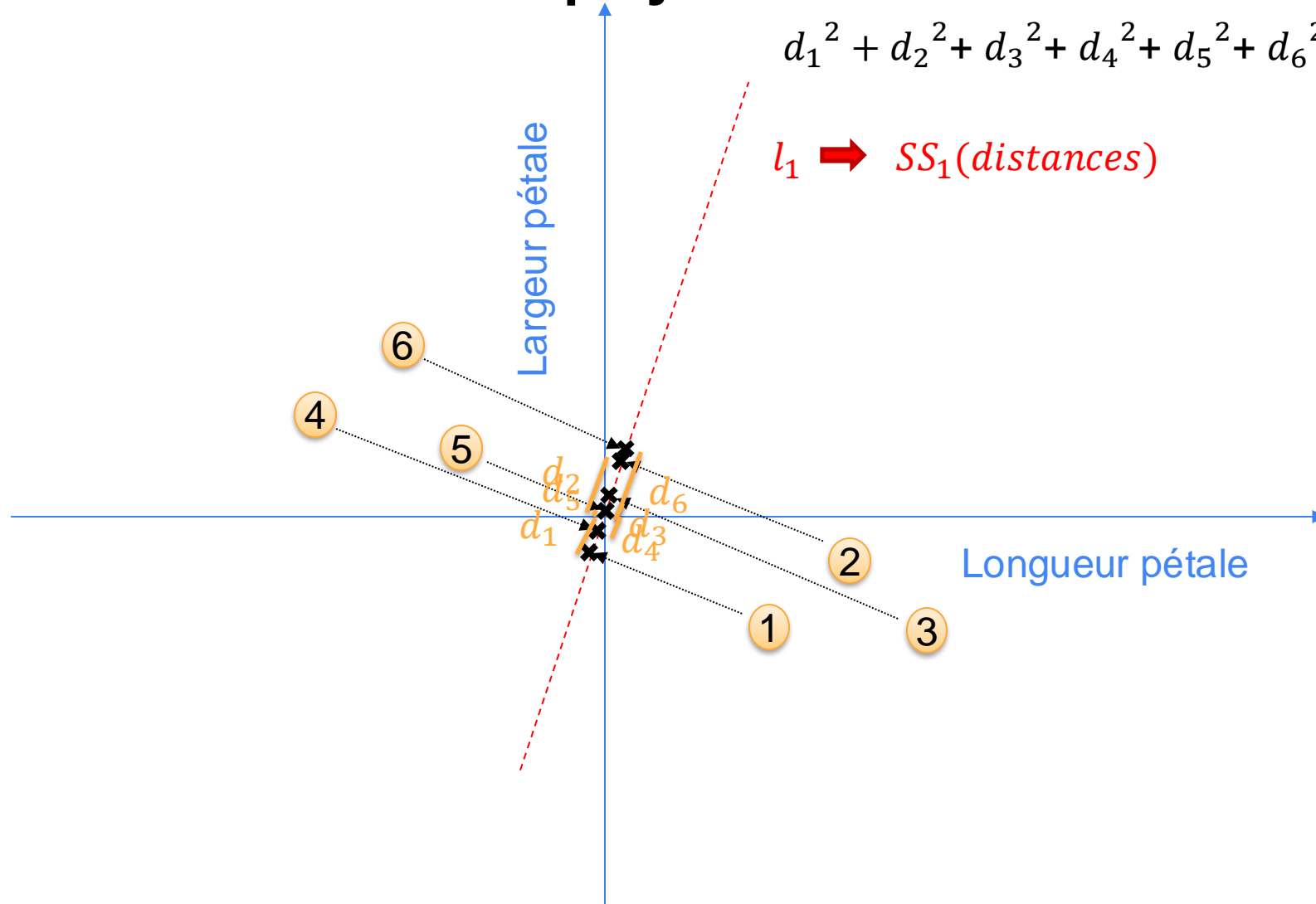
Maximiser la distance entre les projections des observations et l'origine

La distance entre l'observation et l'origine est fixe





Trouver la meilleur projection



$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of square distances}$$

$$= \text{SS}(\text{distances})$$

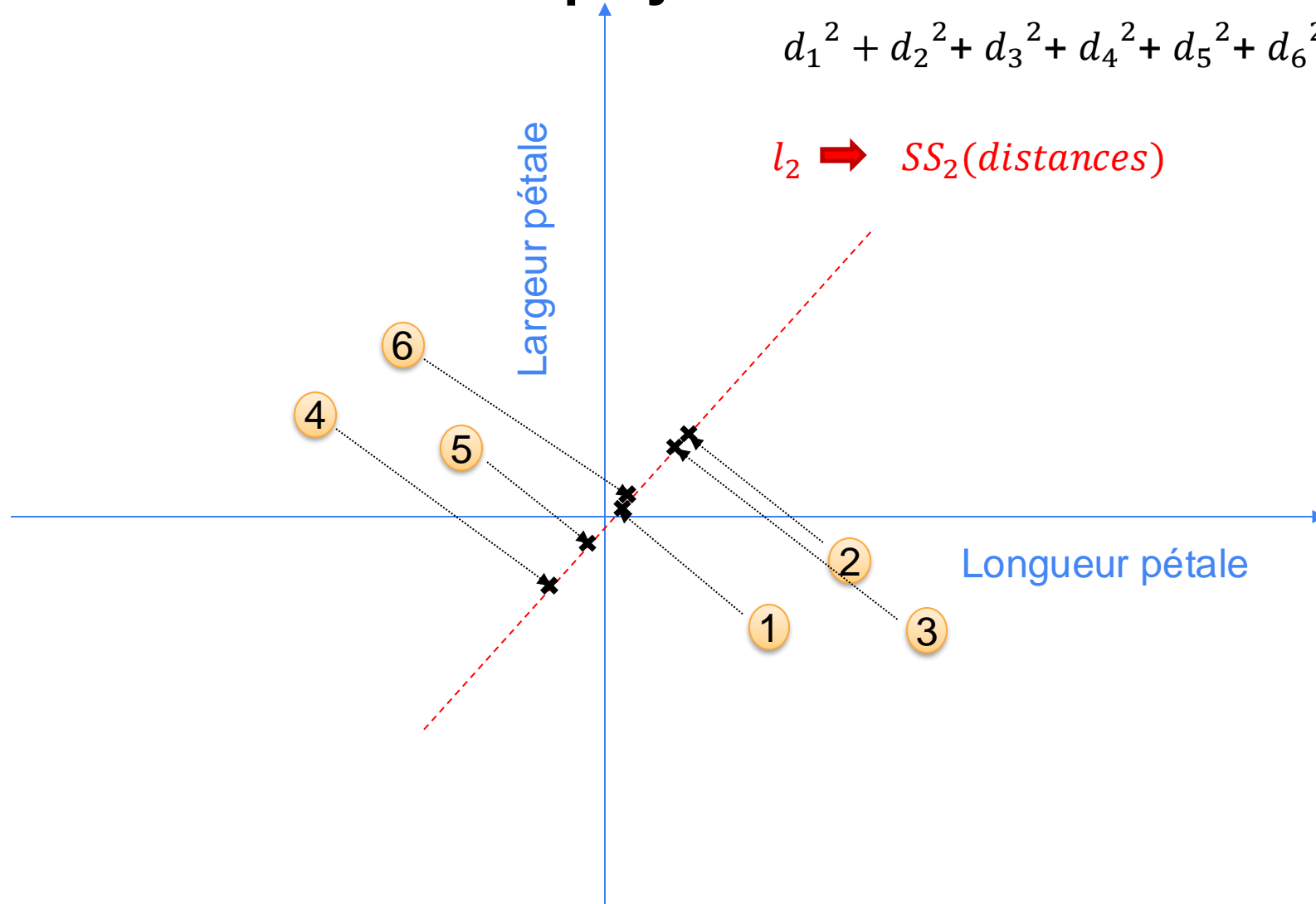
$$l_1 \rightarrow \text{SS}_1(\text{distances})$$



Trouver la meilleur projection

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2$$

$l_2 \Rightarrow SS_2(\text{distances})$

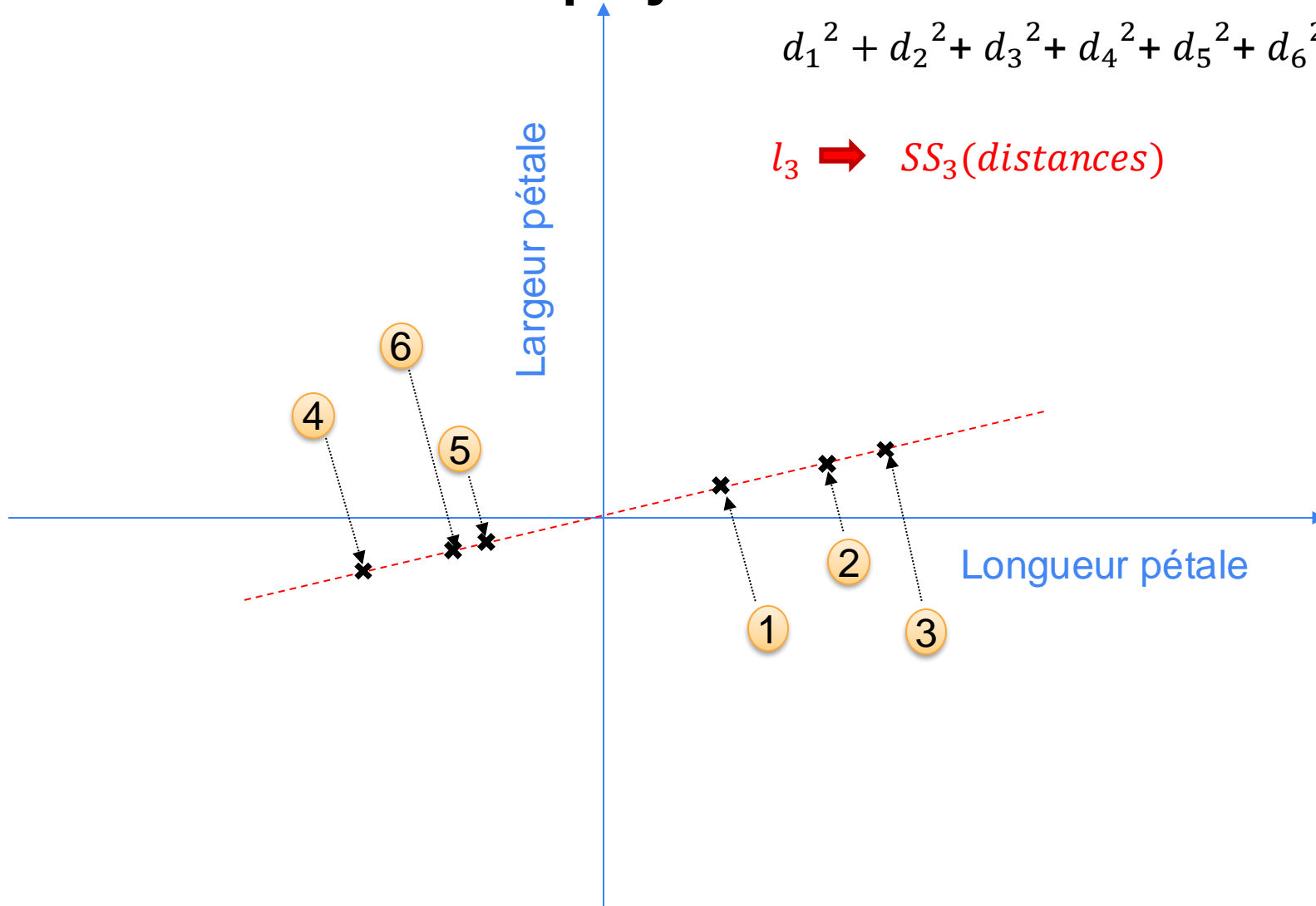




Trouver la meilleur projection

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2$$

$l_3 \Rightarrow SS_3(\text{distances})$

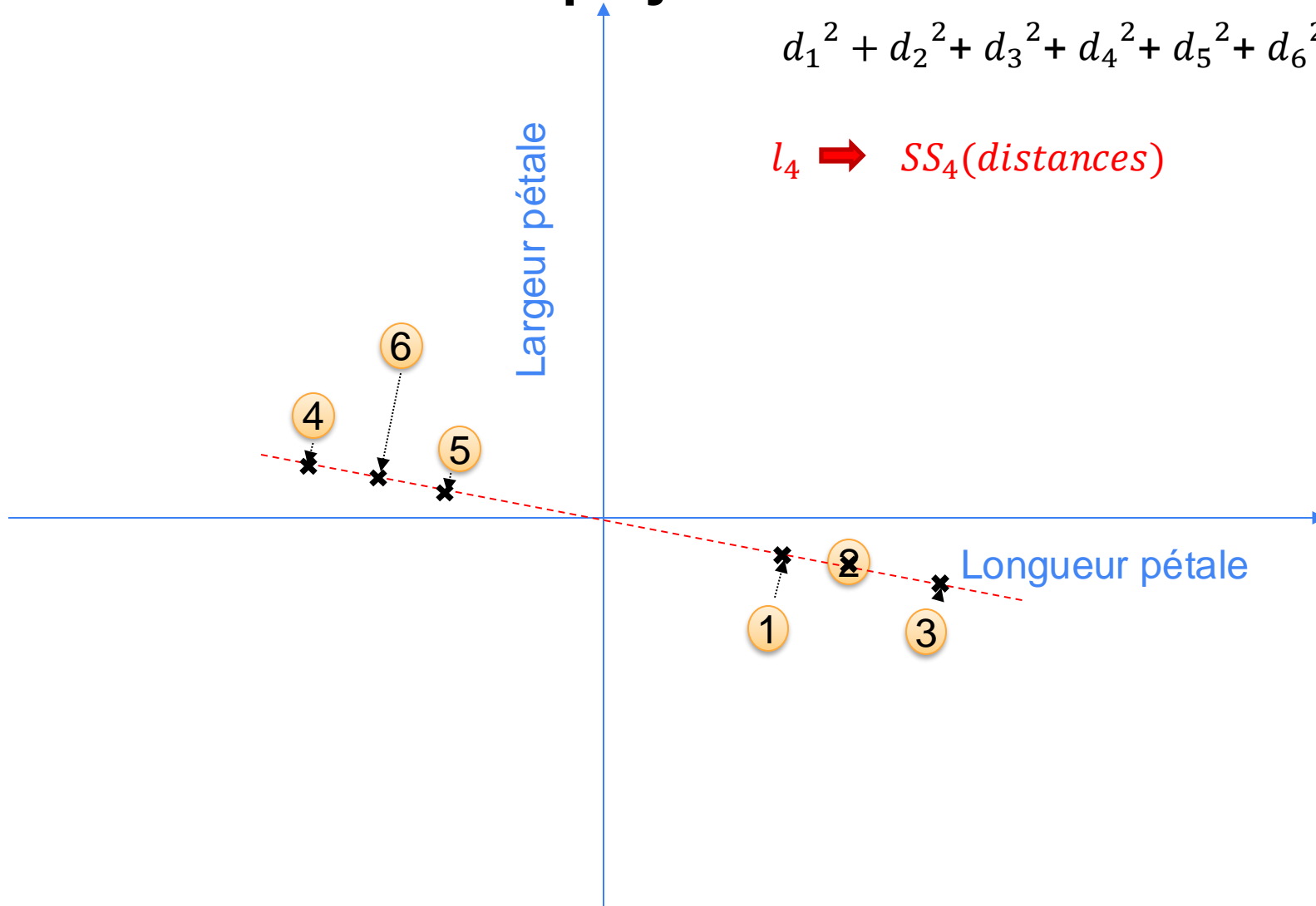




Trouver la meilleur projection

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2$$

$l_4 \Rightarrow SS_4(\text{distances})$



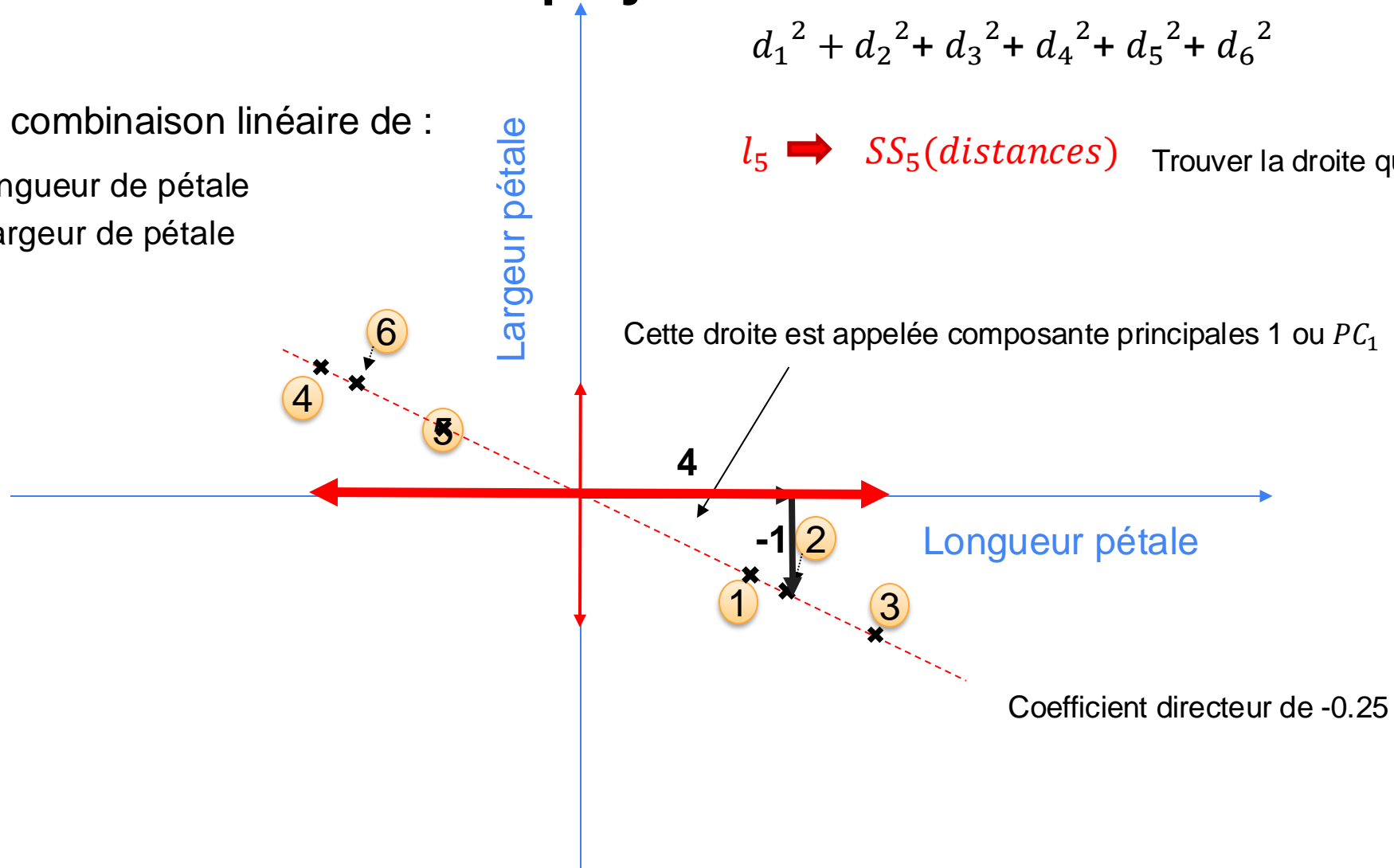


Trouver la meilleur projection

PC_1 est une combinaison linéaire de :

4 part de Longueur de pétale

-1 part de Largeur de pétale

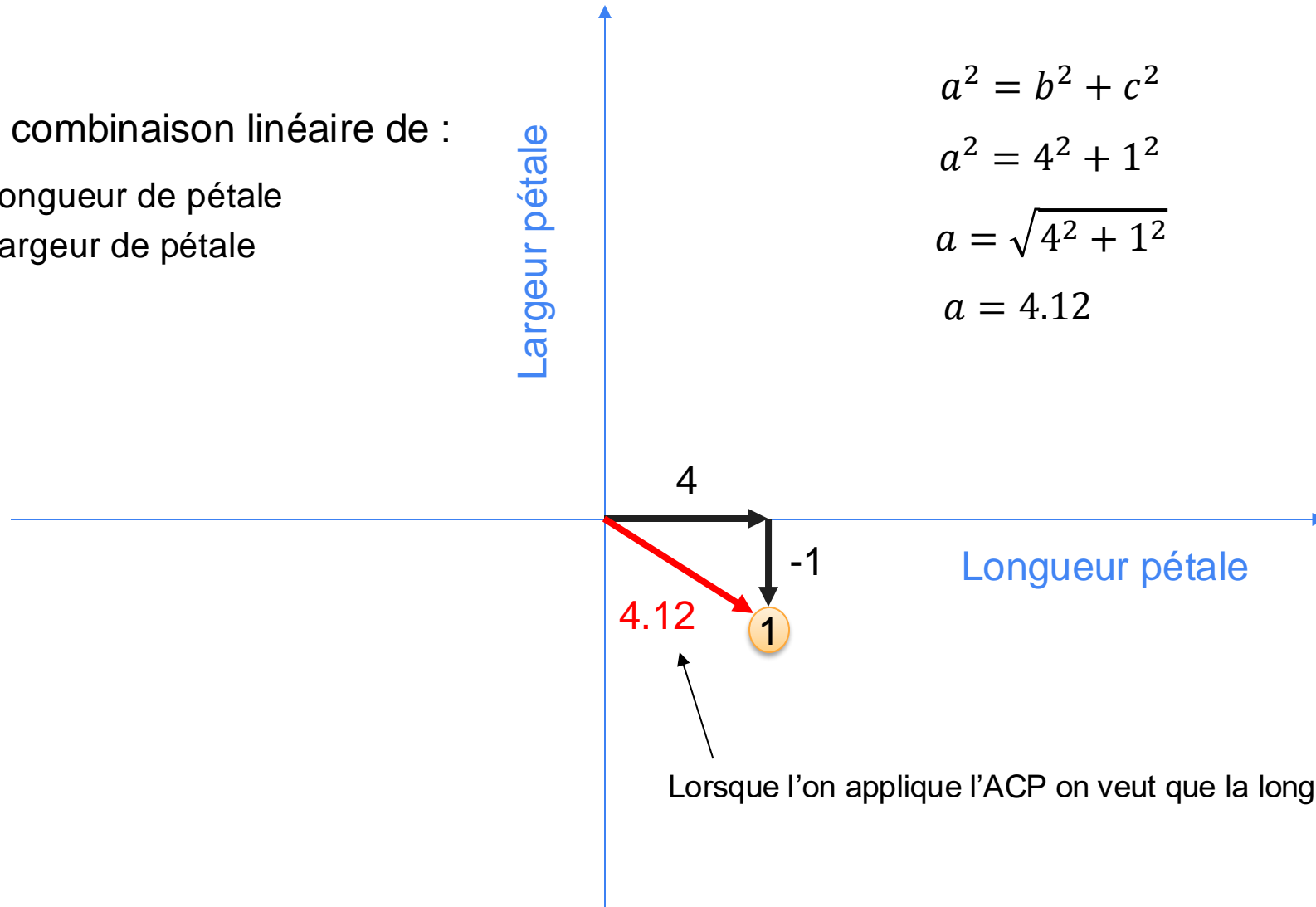




Calcul de PC1

PC_1 est une combinaison linéaire de :

- 4 part de Longueur de pétale
- 1 part de Largeur de pétale



$$a^2 = b^2 + c^2$$

$$a^2 = 4^2 + 1^2$$

$$a = \sqrt{4^2 + 1^2}$$

$$a = 4.12$$

Lorsque l'on applique l'ACP on veut que la longueur de PC_1 soit égale à 1

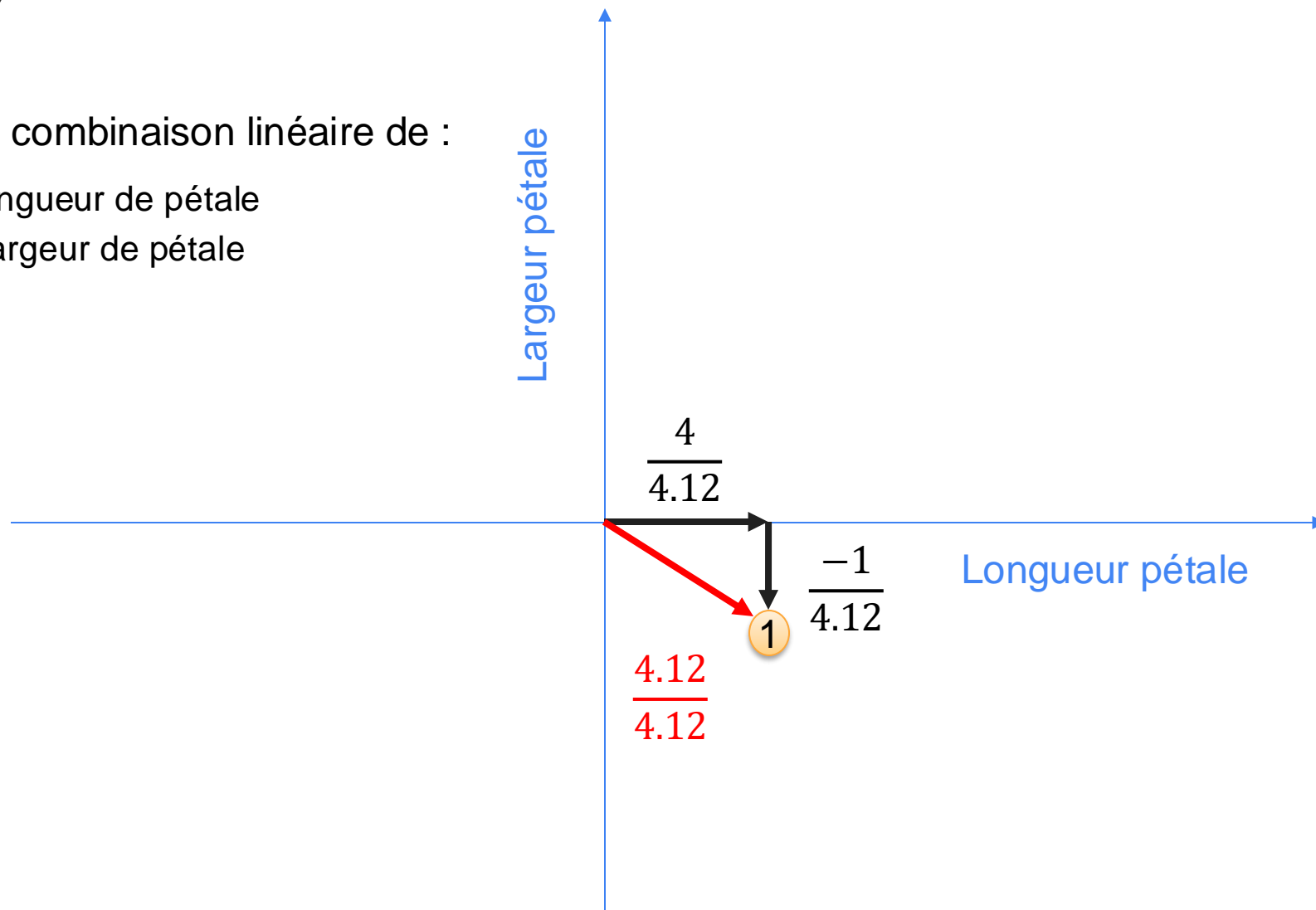


Calcul de PC1

PC_1 est une combinaison linéaire de :

4 part de Longueur de pétale

-1 part de Largeur de pétale





Calcul de PC1

PC_1 est une combinaison linéaire de :

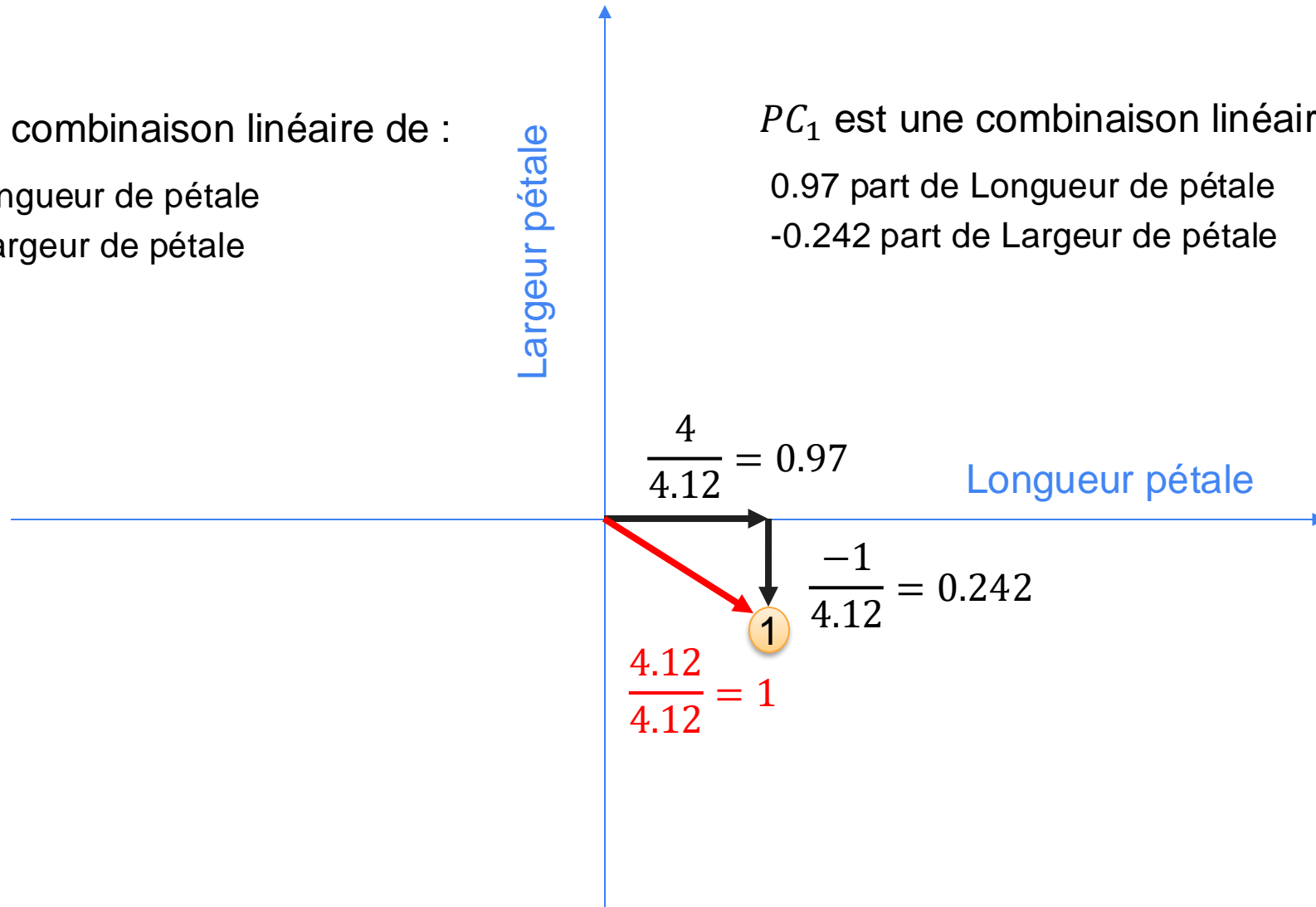
4 part de Longueur de pétale

-1 part de Largeur de pétale

PC_1 est une combinaison linéaire de :

0.97 part de Longueur de pétale

-0.242 part de Largeur de pétale

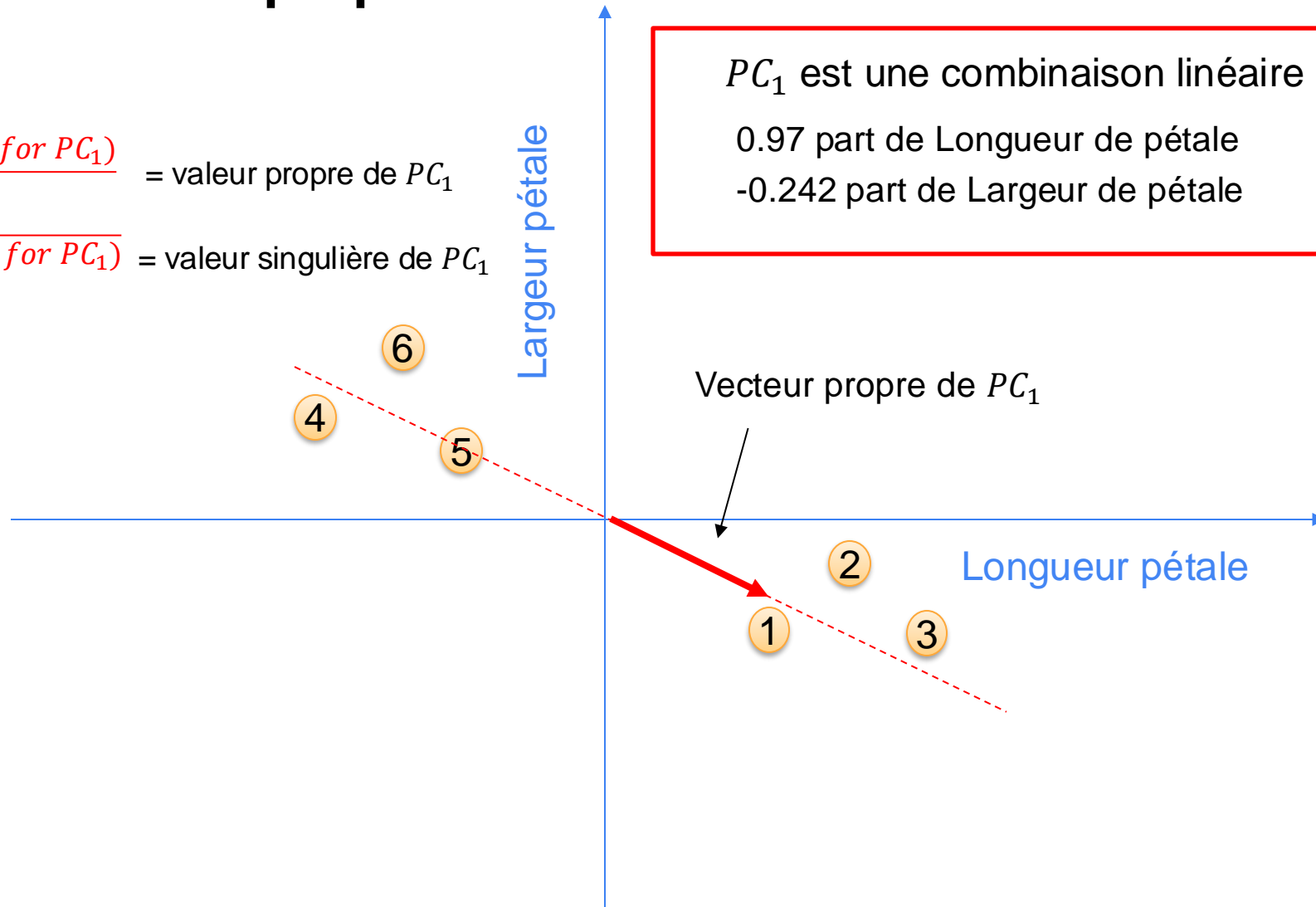




Vecteur propre

$\frac{SS(\text{distances for } PC_1)}{n - 1}$ = valeur propre de PC_1

$\sqrt{SS(\text{distances for } PC_1)}$ = valeur singulière de PC_1



PC_1 est une combinaison linéaire de :

0.97 part de Longueur de pétale

-0.242 part de Largeur de pétale

Loading scores



Comment trouver PC2 ?

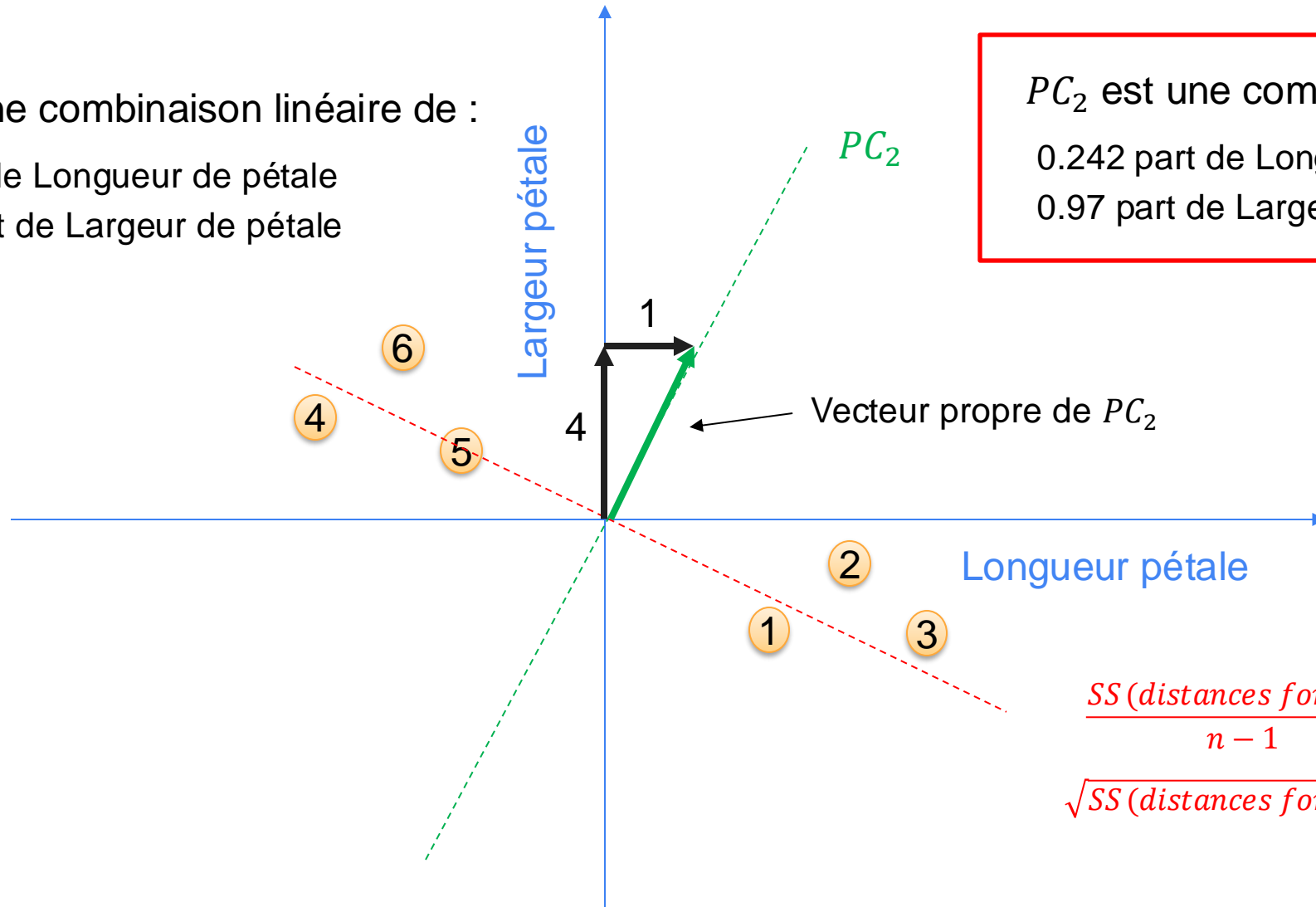
PC_1 est une combinaison linéaire de :

0.97 part de Longueur de pétale
-0.242 part de Largeur de pétale

PC_2 est une combinaison linéaire de :

0.242 part de Longueur de pétale
0.97 part de Largeur de pétale

Loading scores

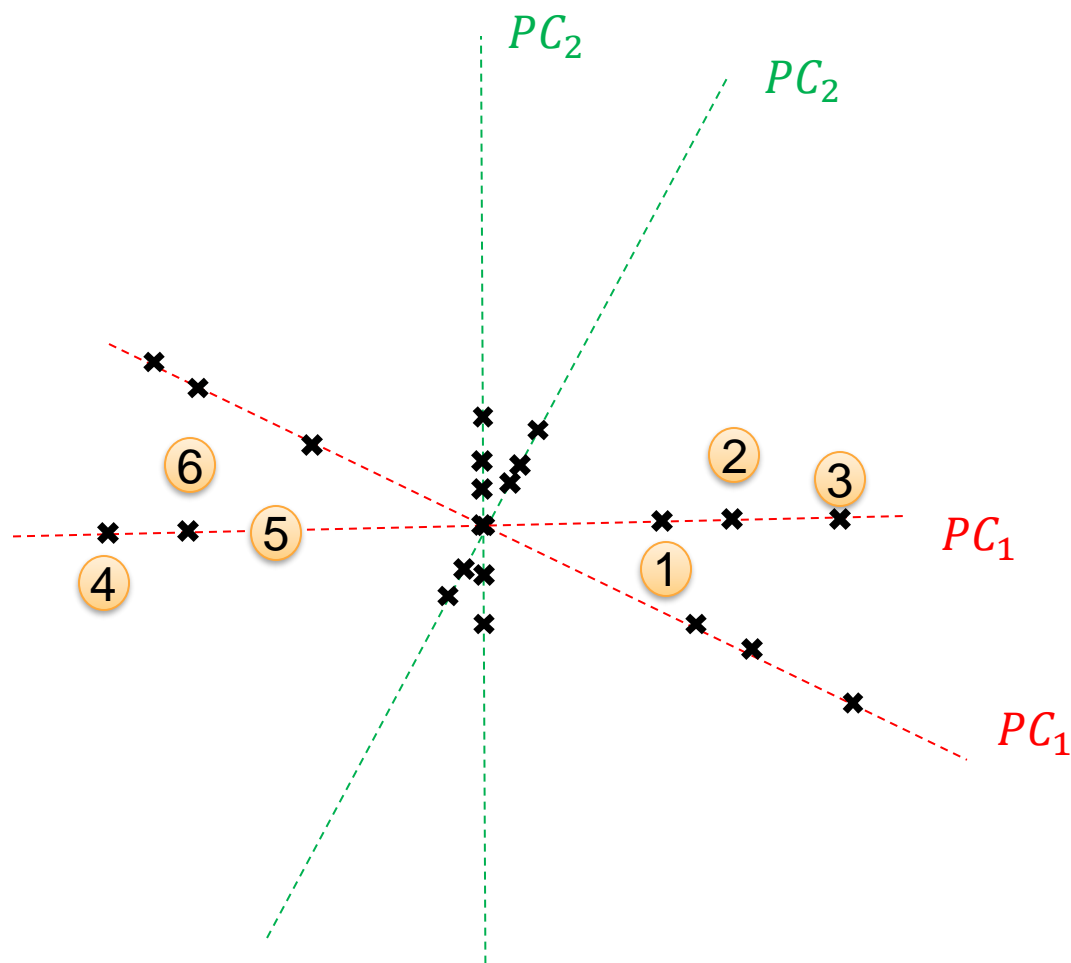


$$\frac{SS(\text{distances for } PC_2)}{n-1} = \text{valeur propre de } PC_2$$

$$\sqrt{SS(\text{distances for } PC_2)} = \text{valeur singulière de } PC_2$$



Comment trouver PC2 ?

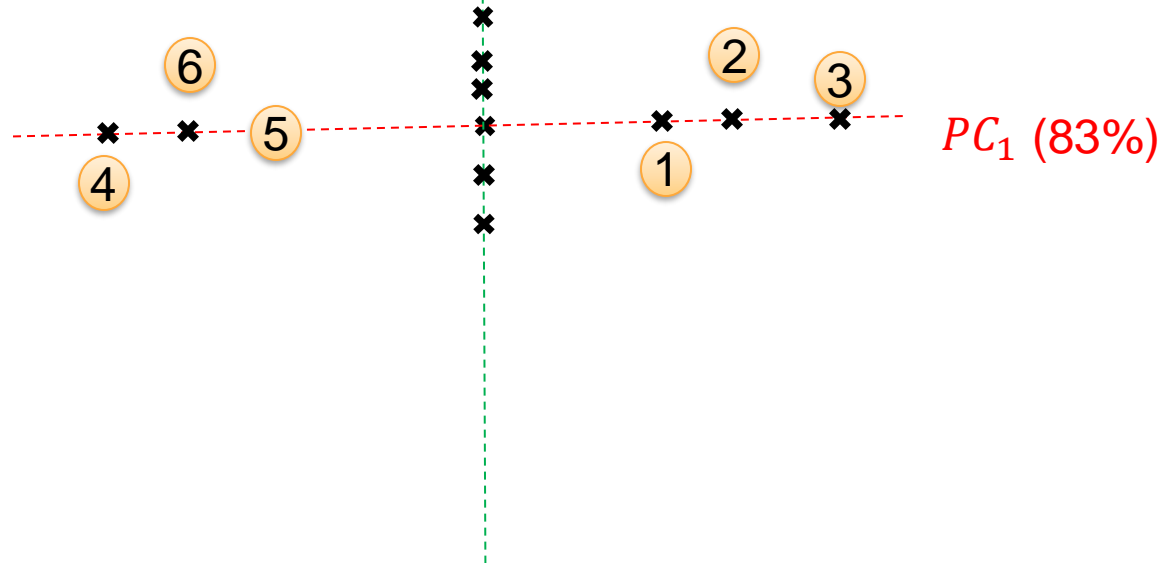




Les valeurs propres

$\frac{SS(\text{distances for } PC_1)}{n-1} = \text{variance propre de } PC_1$

$\frac{SS(\text{distances for } PC_2)}{n-1} = \text{variance propre de } PC_2$



Par exemple :

Variation pour $PC_1=15$ $15/18=0.83$

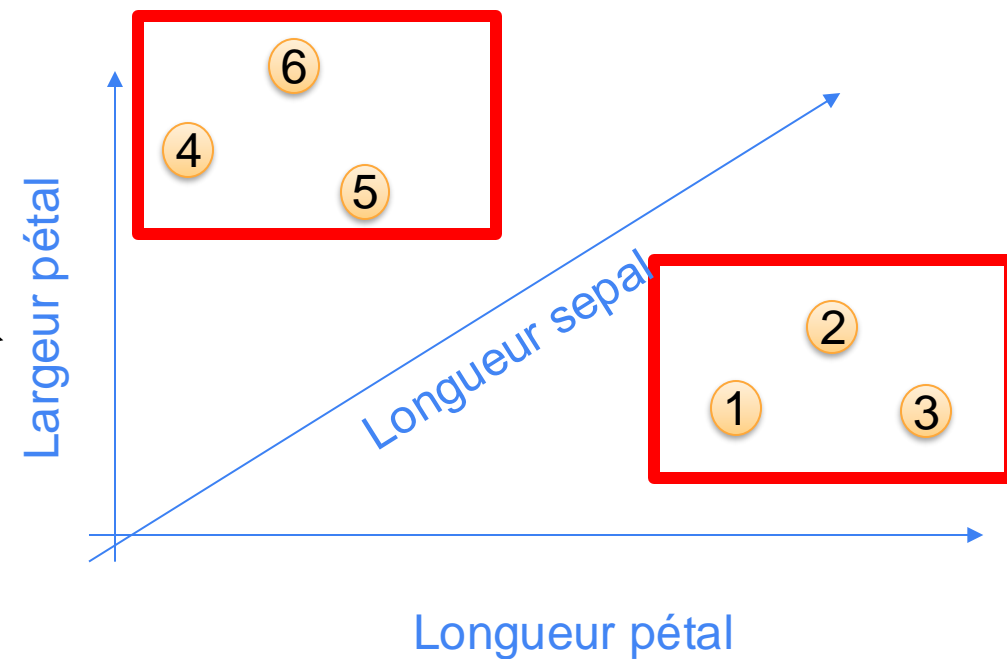
Variation pour $PC_2=3$ $3/18=0.17$

Variation total = $PC_1 + PC_2 = 15 + 3 = 18$



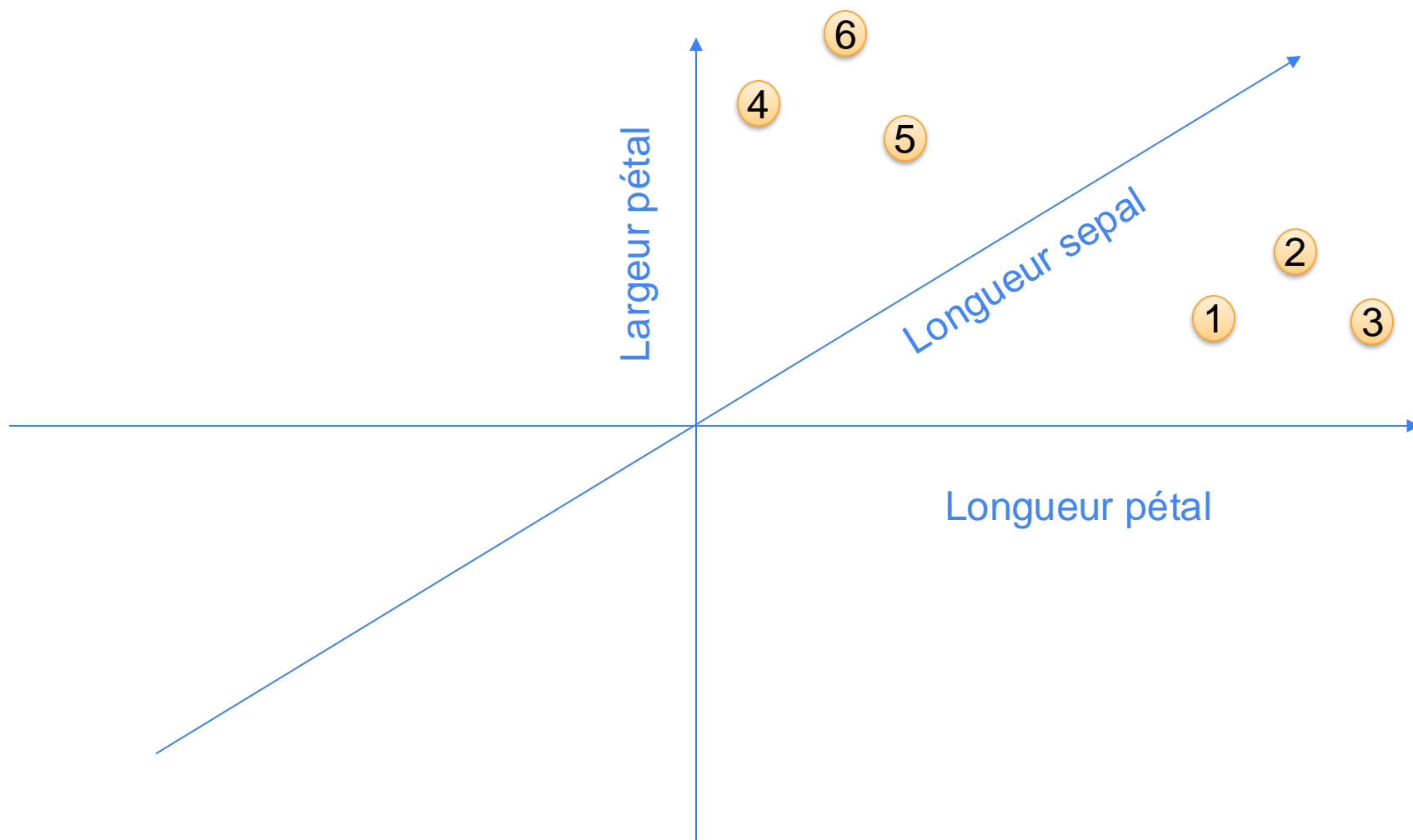
PCA in three dimension

	Longueur pétale	Largeur pétale	Longueur sepal
Fleur_1	11	2	12
Fleur_2	15	5	11
Fleur_3	16	2	14
Fleur_4	3	10	8
Fleur_5	5	9	9
Fleur_6	4	12	0
...





Centrer les données





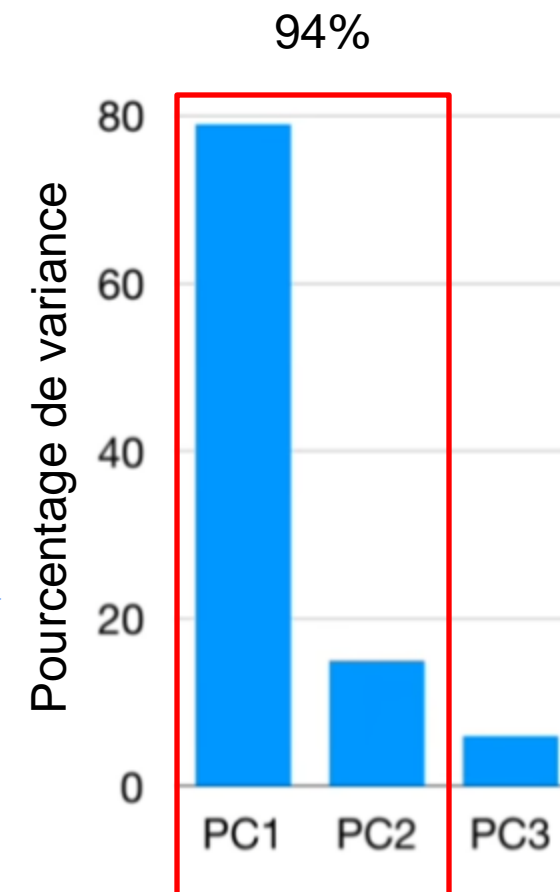
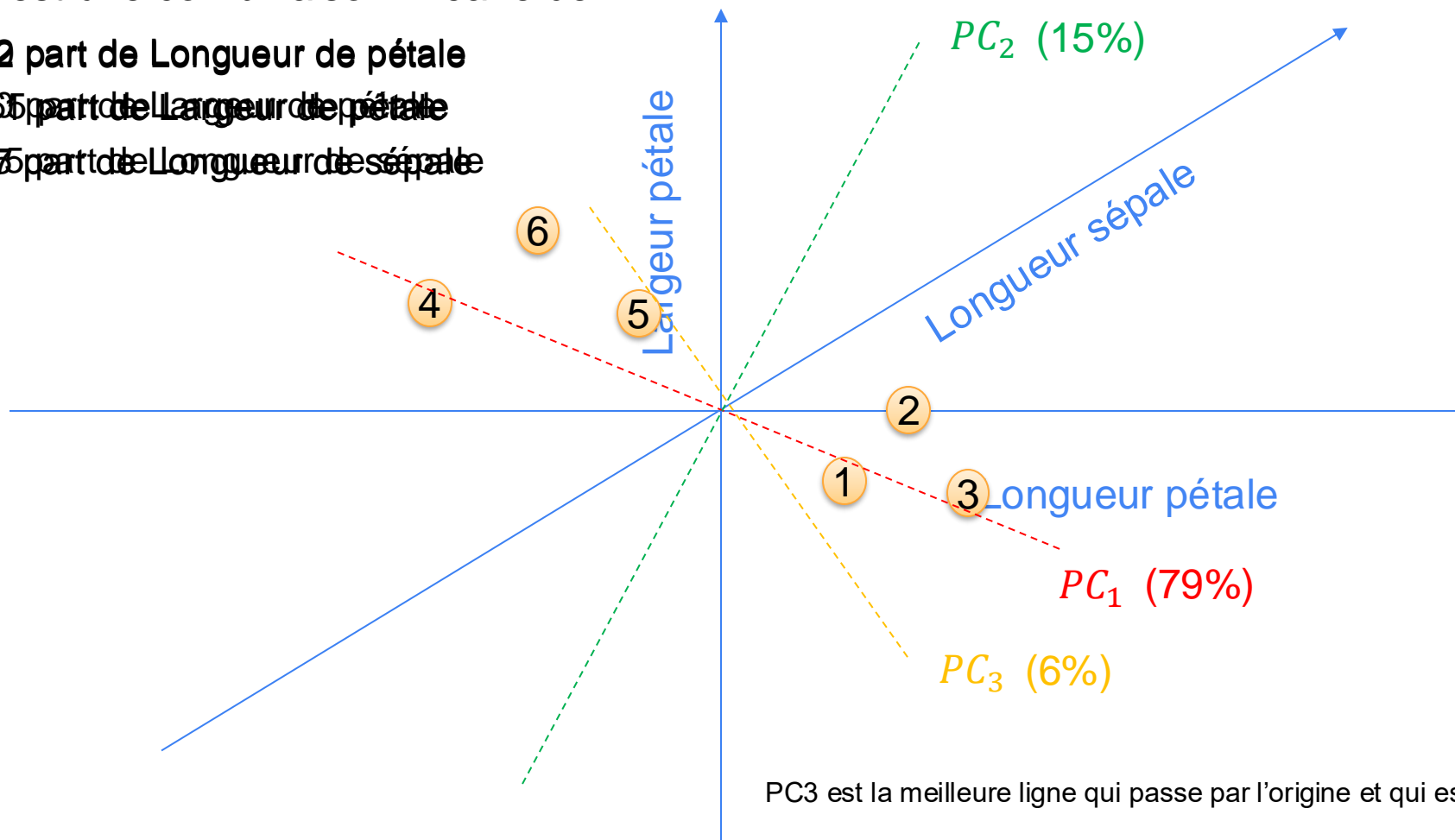
Déterminer les composantes principales

PC_2 est une combinaison linéaire de :

0.82 part de Longueur de pétale

0.56 part de Largeur de pétale

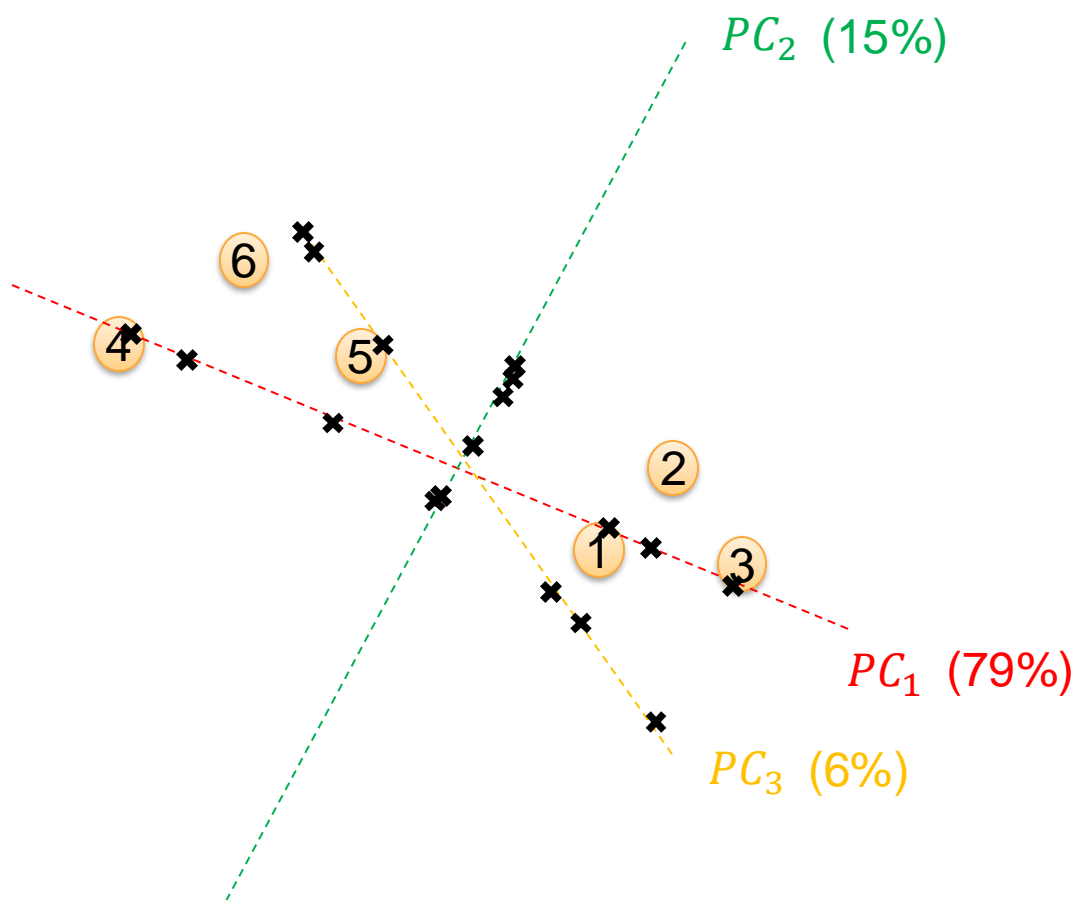
0.75 part de Longueur de sépale



PC_3 est la meilleure ligne qui passe par l'origine et qui est perpendiculaire à PC_1 et PC_2

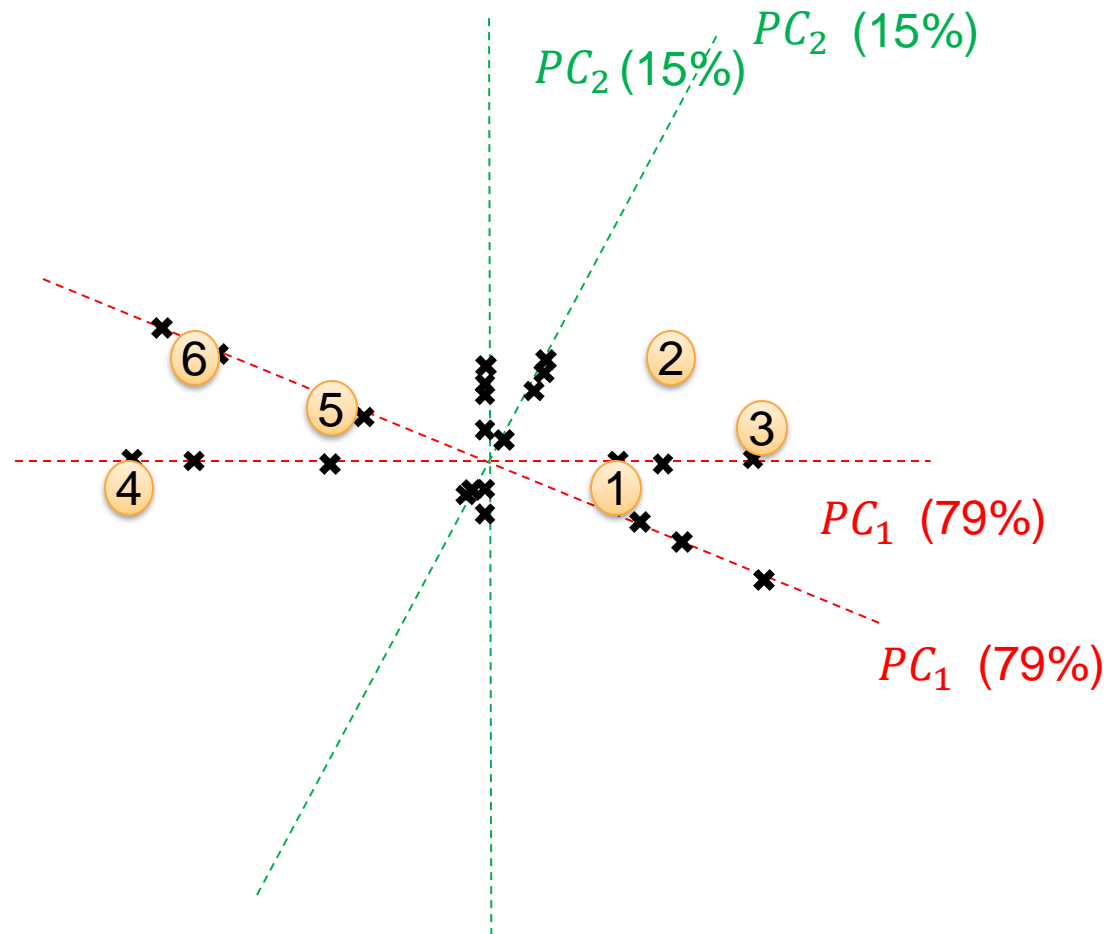


Passer de trois à deux dimensions





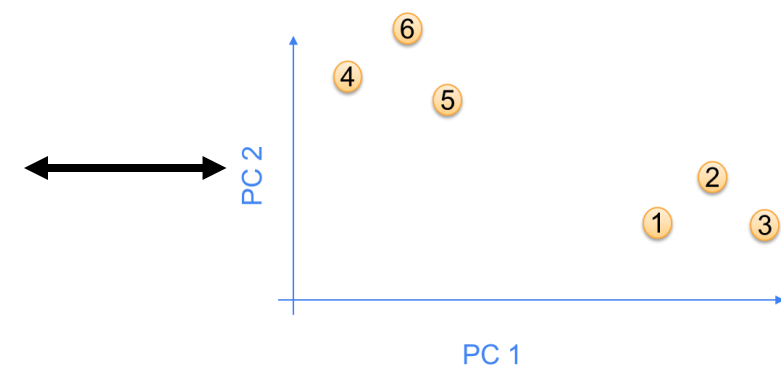
Passer de trois à deux dimensions





De N à deux dimensions

	Longueur pétale	Largeur pétale	Longueur sepal	Largeur sepal	Longueur tige	...
Fleur_1	11	2	12
Fleur_2	15	5	11
Fleur_3	16	2	14
Fleur_4	3	10	8
Fleur_5	5	9	9
Fleur_6	4	12	0
...





De N à deux dimensions

