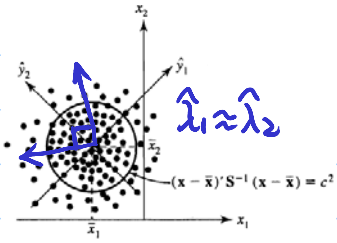
- equal eigenvalues (i.e., $\lambda_i = \lambda_{i+1} = \cdots = \lambda_j$)
  - population principal component
  - sample principal component



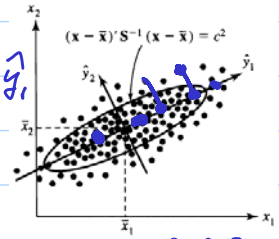
- selecting a subset of variables X_1, …, X_p
  - PCA does not quite reduce the data since PCs are linear combinations of all variables X_1, …, X_p so all variables are in some sense still needed. The technique, however, can help us discard some variables by keeping only those with the highest factor loading

LNp.2-12.

- connection to the singular value decomposition (SVD)
  - Let us apply a SVD to the matrix $\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'$

$$(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')_{n\times p} = U_{n\times p} L_{p\times p} V'_{p\times p}$$

where $U'U = I_p$, $V'V = I_p$, and L is a diagonal matrix. Then,



$Y_k \equiv Y\begin{bmatrix} I_k \\ 0 \end{bmatrix} = [Y_1 | \cdots | Y_k]$ $n\times k$

$\bullet : n\times 3$

$$(n-1)\mathbf{S} = (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')'(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}') = VLU'ULV' = V(LL)V' = VL^2V' \quad X: n\times 3 \ (x_1, x_2, x_3)$$

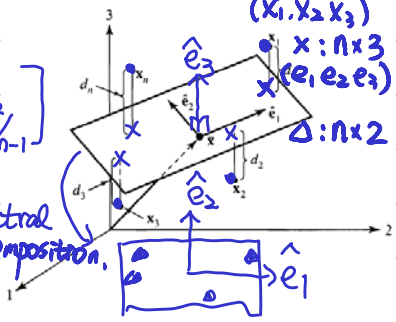
$$\mathbf{S} = V\left(\frac{1}{n-1}L^2\right)V' = V\left(\frac{1}{\sqrt{n-1}}L\right)^2 V' \equiv VL^{*2}V' \quad X: n\times 3 \ (e_1, e_2, e_3)$$

$= \begin{bmatrix} l_1^2/n-1 & 0 \\ 0 & l_p^2/n-1 \end{bmatrix}$ $\Delta: n\times 2$

$= V\Lambda V'$ ← spectral decomposition.

which means

- columns of V: eigenvectors of $\boldsymbol{S}$
- $\frac{l_i^2}{n-1}$: eigenvalues of $\boldsymbol{S}$, where l_i: singular value
- $\mathbf{Y} = (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')V = UL$: PC scores $\cdots \bullet (\hat{e}_1, \hat{e}_2, \hat{e}_3)$



LNp2-18 result 2A.16

- the matrix $\mathbf{B}$ that minimize

$UL\,V'$, $\Lambda = \frac{1}{n-1}L^2 = \begin{bmatrix} \hat{\lambda}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{\lambda}_p \end{bmatrix}$

$$tr[(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}} - \mathbf{B})(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}} - \mathbf{B})'] \quad \text{sum of squared difference} = \sum_{i\,j} (a_{ij} - b_{ij})^2$$

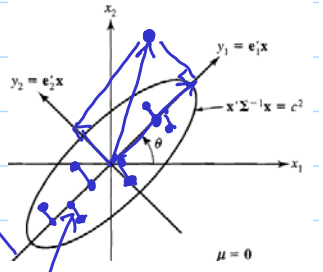
over all $n\times p$ matrices $\mathbf{B}$ having rank no greater than k ($k < p$) is

X_k = $\mathbf{B} = U$ $X: n\times p$ (x_1, x_2, x_3)

$\begin{bmatrix} \hat{\lambda}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{\lambda}_k \end{bmatrix}$ $V' \equiv UL_kV'$, minimum value of sum of squared difference $= \hat{\lambda}_{k+1}^2 + \cdots + \hat{\lambda}_p^2$

- UL_k represents the first k PC scores $[Y_k | 0] \cdots \times(\hat{e}_1, \hat{e}_2, \hat{e}_3)$ $= Y_k^*$

- geometry of PC line (plane)
  - the PC line (plane) minimizes the sum of squared *orthogonal distances* from each data point to the line (plane)



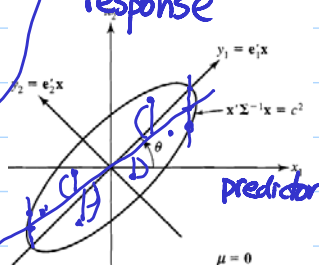
PC line

$$\mathbf{x}_j = (\mathbf{x}_j'\hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1 + (\mathbf{x}_j'\hat{\mathbf{e}}_2)\hat{\mathbf{e}}_2 + \cdots + (\mathbf{x}_j'\hat{\mathbf{e}}_p)\hat{\mathbf{e}}_p$$

PC scores $= \hat{y}_{j1}\hat{\mathbf{e}}_1 + \hat{y}_{j2}\hat{\mathbf{e}}_2 + \cdots + \hat{y}_{jp}\hat{\mathbf{e}}_p$

$= \hat{y}_{j1}\hat{e}_1 + \hat{y}_{j2}\hat{e}_2 + \cdots + \hat{y}_{jk}\hat{e}_k + \hat{y}_{j\,k+1}\hat{e}_{k+1} + \cdots + \hat{y}_{jp}\hat{e}_p$

response



- (c.f.) the least square line in regression minimizes the sum of *vertical distances* from the data point to the line

➢ drawbacks of PCA

Note: PCA does not use the information of other moments (including 1st moment)

- PCA only utilizes information contained in the second moments
- nonlinear structure may be missed
- linear combination of variables may not be meaningful especially if the variables do not represent comparable quantities
- outliers may distort the results

data matrix		dim		cov matrix	sum of variance explained
$X_{n\times p}$	$(\bullet, x_1, x_2, x_3)$	p	$= U L V'$	$V \Lambda V'$	$\hat{\lambda}_1^2 + \cdots + \hat{\lambda}_p^2$
X_{k} $n\times p$	$(\times, x_1, x_2, x_3)$	k	$= U_k L_k V'$	$V \Lambda_k V'$	$\hat{\lambda}_1^2 + \cdots + \hat{\lambda}_k^2$
Y $n\times p$	$(\bullet, \hat{e}_1, \hat{e}_2, \hat{e}_3)$	p	$= UL$	Λ	$\hat{\lambda}_1^2 + \cdots + \hat{\lambda}_p^2$
Y_k^* $n\times p$	$(\times, \hat{e}_1, \hat{e}_2, \hat{e}_3)$	k	$= U L_k$	Λ_k	$\hat{\lambda}_1^2 + \cdots + \hat{\lambda}_k^2$
Y_k $n\times k$	$(\Delta, \hat{e}_1, \hat{e}_2)$ $\uparrow \begin{bmatrix}1\\0\end{bmatrix}$ $\uparrow \begin{bmatrix}0\\1\end{bmatrix}$	k	$= UL \times \begin{bmatrix} I_k \\ 0 \end{bmatrix}$	Λ_k	$\hat{\lambda}_1^2 + \cdots + \hat{\lambda}_k^2$

❖ **Reading**: Textbook, 8.1, 8.2, 8.3, 8.4, 8.5, Supplement 8A