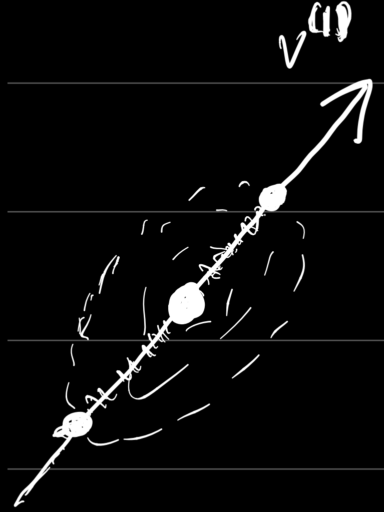
Conclusion: the direction with largest data variance is determined by an eigenvector associated to the largest eigenvalue of $C = \frac{1}{n} XX^T$ or equivalently of XX^T.



To find the direction with the second largest data variance, say $v^{(2)}$, we require this direction to be orthogonal to $v^{(1)}$

Recall: projected data variance along $v \in S^{d-1}$

$$= v^T C v = \frac{1}{n} v^T (XX^T) v$$

Thus

$$v^{(2)} = \underset{\substack{\|v\|=1 \\ v \perp v^{(1)}}}{\arg\max}\ v^T C v$$

From the previous proposition, $v^{(2)}$ is an eigenvector associated to the second largest eigenvalue of C, or equivalently of XX^T.

Continue as above, the direction that has the third largest data variance and is orthogonal to $v^{(1)}$, $v^{(2)}$ is determined by an eigenvector associated to the third largest eigenvalue of C or XX^T

$\vdots$

Thm: The d-dimensional subspace with the largest projected data variance is spanned by the d eigenvectors associated to the d largest eigenvalues of C or XX^T.

Remarks: (1) If $X = U \Sigma V^T$ is the SVD, then

$$XX^T = U\Sigma \underbrace{V^T V}_{=I} \Sigma^T U^T = U \underbrace{\Sigma\Sigma^T}_{\text{diagonal}} U^T$$

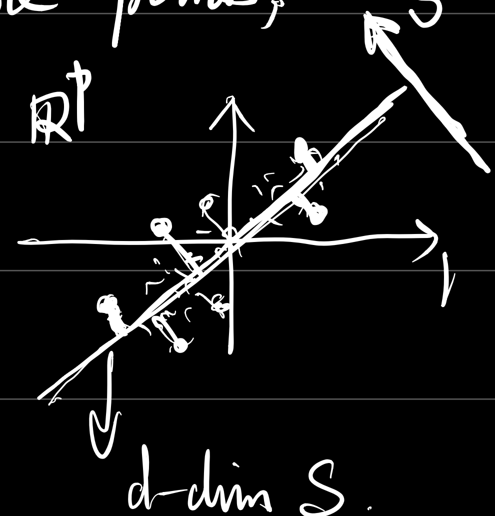
is the eigenvalue decomposition of XX^T, so the eigenvectors associated to the d largest eigenvalues of XX^T are the first d columns of $U = \begin{pmatrix} | & & | \\ u^{(1)} & \cdots & u^{(p)} \\ | & & | \end{pmatrix}$

(2) Given a sample point x, the <u>d-principal component of x</u> is defined as

$$\mathrm{Proj}_{\mathrm{span}\{u^{(1)},\dots,u^{(d)}\}} x = \underbrace{(x^T u^{(1)})}_{\text{scalar projection of } x \text{ on } u^{(1)}} \overset{\text{assigns the direction of } u^{(1)}}{u^{(1)}} + \cdots + (x^T u^{(d)}) u^{(d)}$$

3.2 PCA: Geometric Viewpoint.

Q: Given centered sample points $x^{(1)}, \dots, x^{(n)} \in \mathbb{R}^p$, find a d-dimensional subspace $S \subset \mathbb{R}^p$ that "best fit" the sample points, in the sense that the distances from the sample points to S are minimized.



To compute the distances, let $\{u^{(1)}, \dots, u^{(d)}\}$ be an orthonormal basis of S, and $\{u^{(d+1)}, \dots, u^{(p)}\}$ of $S^\perp$ be an orthonormal basis of $S^\perp$.

then $\{u^{(1)}, \dots, u^{(p)}\}$ is an orthonormal basis of $\mathbb{R}^p$.

For a sample point $x^{(i)}$, we have

$$x^{(i)} = \underbrace{\left(x^{(i)T} u^{(1)}\right) u^{(1)} + \dots + \left(x^{(i)T} u^{(d)}\right) u^{(d)}}_{\in S} + \underbrace{\dots + \left(x^{(i)T} u^{(p)}\right) u^{(p)}}_{\in S^\perp}$$

squared distance from $x^{(i)}$ to S

$$= \| (x^{(i)T} u^{(d+1)}) u^{(d+1)} + \cdots + (x^{(i)T} u^{(p)}) u^{(p)} \|^2$$