Obviously, $U\frac{1}{(n-1)}D^2U^\top$ is a symmetric matrix whose eigenvalues are $\frac{\sigma_1^2}{n-1} \geq \cdots \geq \frac{\sigma_d^2}{n-1}$, and the columns of U form an orthonormal basis of unit eigenvectors.

We proceed by induction on k. For the base case, $k = 1$, maximizing $\mathrm{var}(Y)$ is equivalent to maximizing

$$v^\top U\frac{1}{(n-1)}D^2U^\top v,$$

where v is a unit vector. By Proposition 23.10, the maximum of the above quantity is the largest eigenvalue of $U\frac{1}{(n-1)}D^2U^\top$, namely $\frac{\sigma_1^2}{n-1}$, and it is achieved for u_1, the first columnn of U. Now we get

$$Y_1 = (X - \mu)u_1 = VDU^\top u_1,$$

and since the columns of U form an orthonormal basis, $U^\top u_1 = e_1 = (1, 0, \ldots, 0)$, and so Y_1 is indeed the first column of VD.

By the induction hypothesis, the centered points $Y_1, \ldots, Y_k$, where $Y_h = (X - \mu)u_h$ and $u_1, \ldots, u_k$ are the first k columns of U, are k principal components of X. Because

$$\mathrm{cov}(Y, Z) = v^\top U\frac{1}{(n-1)}D^2U^\top w,$$

where $Y = (X - \mu)v$ and $Z = (X - \mu)w$, the condition $\mathrm{cov}(Y_h, Z) = 0$ for $h = 1, \ldots, k$ is equivalent to the fact that w belongs to the orthogonal complement of the subspace spanned by $\{u_1, \ldots, u_k\}$, and maximizing $\mathrm{var}(Z)$ subject to $\mathrm{cov}(Y_h, Z) = 0$ for $h = 1, \ldots, k$ is equivalent to maximizing

$$w^\top U\frac{1}{(n-1)}D^2U^\top w,$$

where w is a unit vector orthogonal to the subspace spanned by $\{u_1, \ldots, u_k\}$. By Proposition 23.10, the maximum of the above quantity is the $(k+1)$th eigenvalue of $U\frac{1}{(n-1)}D^2U^\top$, namely $\frac{\sigma_{k+1}^2}{n-1}$, and it is achieved for u_{k+1}, the $(k+1)$th columnn of U. Now we get

$$Y_{k+1} = (X - \mu)u_{k+1} = VDU^\top u_{k+1},$$

and since the columns of U form an orthonormal basis, $U^\top u_{k+1} = e_{k+1}$, and Y_{k+1} is indeed the $(k+1)$th column of VD, which completes the proof of the induction step. $\square$

The d columns $u_1, \ldots, u_d$ of U are usually called the *principal directions* of $X - \mu$ (and X). We note that not only do we have $\mathrm{cov}(Y_h, Y_k) = 0$ whenever $h \neq k$, but the directions $u_1, \ldots, u_d$ along which the data are projected are mutually orthogonal.

Example 23.10. For the centered data set of our bearded mathematicians (Example 23.9) we have $X - \mu = V\Sigma U^\top$, where Σ has two nonzero singular values, $\sigma_1 = 116.9803, \sigma_2 = 21.7812$, and with

$$U = \begin{pmatrix} 0.9995 & 0.0325 \\ 0.0325 & -0.9995 \end{pmatrix},$$