The above suggests that we should move the origin to the centroid μ of the X_i's and consider the matrix $X - \mu$ of the centered data points $X_i - \mu$.

From now on beware that we denote the columns of $X - \mu$ by $C_1, \ldots, C_d$ and that Y denotes the *centered* point $Y = (X - \mu)v = \sum_{j=1}^{d} v_j C_j$, where v is a unit vector.

Basic idea of PCA: The principal components of X are *uncorrelated* projections Y of the data points X_1, …, X_n onto some directions v (where the v's are unit vectors) such that $\mathrm{var}(Y)$ is maximal.

This suggests the following definition:

Definition 23.2. Given an $n \times d$ matrix X of data points $X_1, \ldots, X_n$, if μ is the centroid of the X_i's, then a *first principal component of* X *(first PC)* is a centered point $Y_1 = (X - \mu)v_1$, the projection of $X_1, \ldots, X_n$ onto a direction v_1 such that $\mathrm{var}(Y_1)$ is maximized, where v_1 is a unit vector (recall that $Y_1 = (X - \mu)v_1$ is a linear combination of the C_j's, the columns of $X - \mu$).

More generally, if $Y_1, \ldots, Y_k$ are k principal components of X along some unit vectors $v_1, \ldots, v_k$, where $1 \leq k < d$, a $(k+1)th$ *principal component of* X *($(k+1)th$ PC)* is a centered point $Y_{k+1} = (X - \mu)v_{k+1}$, the projection of $X_1, \ldots, X_n$ onto some direction v_{k+1} such that $\mathrm{var}(Y_{k+1})$ is maximized, subject to $\mathrm{cov}(Y_h, Y_{k+1}) = 0$ for all h with $1 \leq h \leq k$, and where v_{k+1} is a unit vector (recall that $Y_h = (X - \mu)v_h$ is a linear combination of the C_j's). The v_h are called *principal directions*.

The following proposition is the key to the main result about PCA. This result was already proven in Proposition 17.23 except that the eigenvalues were listed in increasing order. For the reader's convenience we prove it again.

Proposition 23.10. *If A is a symmetric $d \times d$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and if $(u_1, \ldots, u_d)$ is any orthonormal basis of eigenvectors of A, where u_i is a unit eigenvector associated with λ_i, then*

$$\max_{x \neq 0} \frac{x^\top A x}{x^\top x} = \lambda_1$$

(with the maximum attained for $x = u_1$) and

$$\max_{x \neq 0, x \in \{u_1, \ldots, u_k\}^\perp} \frac{x^\top A x}{x^\top x} = \lambda_{k+1}$$

(with the maximum attained for $x = u_{k+1}$), where $1 \leq k \leq d - 1$.

Proof. First observe that

$$\max_{x \neq 0} \frac{x^\top A x}{x^\top x} = \max_{x} \{x^\top A x \mid x^\top x = 1\},$$