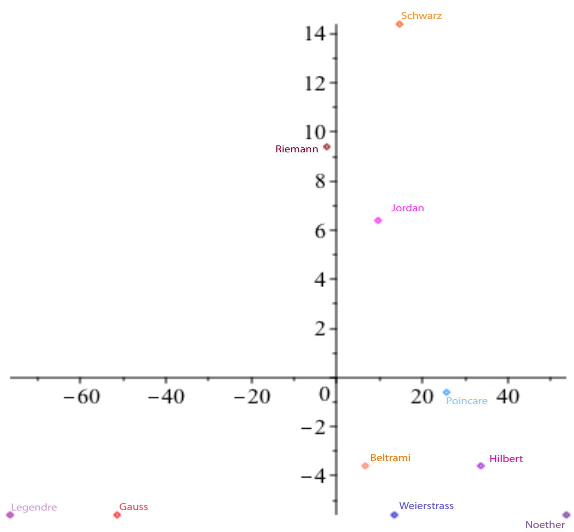

Figure 23.3: The centered data points of Example 23.9.

Therefore, the vector $Y \in \mathbb{R}^n$ consisting of the coordinates of the projections of $X_1, \ldots, X_n$ onto the line spanned by v is given by $Y = Xv$, and this is the linear combination

$$Xv = v_1 C_1 + \cdots + v_d C_d$$

of the columns of X (with $v = (v_1, \ldots, v_d)$).

Observe that because μ_j is the mean of the vector C_j (the jth column of X), we get

$$\overline{Y} = \overline{Xv} = v_1\mu_1 + \cdots + v_d\mu_d,$$

and so the centered point $Y - \overline{Y}$ is given by

$$Y - \overline{Y} = v_1(C_1 - \mu_1) + \cdots + v_d(C_d - \mu_d) = (X - \mu)v.$$

Furthermore, if $Y = Xv$ and $Z = Xw$, then

$$\begin{aligned} \mathrm{cov}(Y, Z) &= \frac{((X-\mu)v)^\top (X-\mu)w}{n-1} \\ &= v^\top \frac{1}{n-1}(X-\mu)^\top (X-\mu)w \\ &= v^\top \Sigma w, \end{aligned}$$

where Σ is the covariance matrix of X. Since $Y - \overline{Y}$ has zero mean, we have

$$\mathrm{var}(Y) = \mathrm{var}(Y - \overline{Y}) = v^\top \frac{1}{n-1}(X-\mu)^\top (X-\mu)v.$$