Recognizing that $||\mathbf{e}|| = 1$, partially differentiating with respect to a_k, and setting the derivative to zero, we obtain

$$a_k = \mathbf{e}^t(\mathbf{x}_k - \mathbf{m}). \tag{83}$$

Geometrically, this result merely says that we obtain a least-squares solution by projecting the vector $\mathbf{x}_k$ onto the line in the direction of $\mathbf{e}$ that passes through the sample mean.

SCATTER MATRIX

This brings us to the more interesting problem of finding the *best* direction $\mathbf{e}$ for the line. The solution to this problem involves the so-called *scatter matrix* $\mathbf{S}$ defined by

$$\mathbf{S} = \sum_{k=1}^{n}(\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^t. \tag{84}$$

The scatter matrix should look familiar—it is merely $n - 1$ times the sample covariance matrix. It arises here when we substitute a_k found in Eq. 83 into Eq. 82 to obtain

$$\begin{aligned} J_1(\mathbf{e}) &= \sum_{k=1}^{n} a_k^2 - 2\sum_{k=1}^{n} a_k^2 + \sum_{k=1}^{n} ||\mathbf{x}_k - \mathbf{m}||^2 \\ &= -\sum_{k=1}^{n}[\mathbf{e}^t(\mathbf{x}_k - \mathbf{m})]^2 + \sum_{k=1}^{n} ||\mathbf{x}_k - \mathbf{m}||^2 \\ &= -\sum_{k=1}^{n} \mathbf{e}^t(\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^t\mathbf{e} + \sum_{k=1}^{n} ||\mathbf{x}_k - \mathbf{m}||^2 \\ &= -\mathbf{e}^t\mathbf{S}\mathbf{e} + \sum_{k=1}^{n} ||\mathbf{x}_k - \mathbf{m}||^2. \end{aligned} \tag{85}$$

Clearly, the vector $\mathbf{e}$ that minimizes J_1 also maximizes $\mathbf{e}^t\mathbf{S}\mathbf{e}$. We use the method of Lagrange multipliers (described in Section A.3 of the Appendix) to maximize $\mathbf{e}^t\mathbf{S}\mathbf{e}$ subject to the constraint that $||\mathbf{e}|| = 1$. Letting λ be the undetermined multiplier, we differentiate

$$u = \mathbf{e}^t\mathbf{S}\mathbf{e} - \lambda(\mathbf{e}^t\mathbf{e} - 1) \tag{86}$$

with respect to $\mathbf{e}$ to obtain

$$\frac{\partial u}{\partial \mathbf{e}} = 2\mathbf{S}\mathbf{e} - 2\lambda\mathbf{e}. \tag{87}$$

Setting this gradient vector equal to zero, we see that $\mathbf{e}$ must be an eigenvector of the scatter matrix:

$$\mathbf{S}\mathbf{e} = \lambda\mathbf{e}. \tag{88}$$

In particular, because $\mathbf{e}^t\mathbf{S}\mathbf{e} = \lambda\mathbf{e}^t\mathbf{e} = \lambda$, it follows that to maximize $\mathbf{e}^t\mathbf{S}\mathbf{e}$, we want to select the eigenvector corresponding to the largest eigenvalue of the scatter matrix. In other words, to find the best one-dimensional projection of the data (best in the least-