we can write this in the form

$$E^{\text{new}} - E^{\text{old}} = -\sum_n \ln \left\{ \frac{\sum_j P^{\text{new}}(j) p^{\text{new}}(\mathbf{x}^n|j)}{p^{\text{old}}(\mathbf{x}^n)} \frac{P^{\text{old}}(j|\mathbf{x}^n)}{P^{\text{old}}(j|\mathbf{x}^n)} \right\} \tag{2.89}$$

where the last factor inside the brackets is simply the identity. We now make use of Jensen's inequality (Exercise 2.13) which says that, given a set of numbers $\lambda_j \geq 0$ such that $\sum_j \lambda_j = 1$,

$$\ln \left(\sum_j \lambda_j x_j \right) \geq \sum_j \lambda_j \ln(x_j). \tag{2.90}$$

Since the probabilities $P^{\text{old}}(j|\mathbf{x})$ in the numerator of (2.89) sum to unity, they can play the role of the λ_j in (2.90). This gives

$$E^{\text{new}} - E^{\text{old}} \leq -\sum_n \sum_j P^{\text{old}}(j|\mathbf{x}^n) \ln \left\{ \frac{P^{\text{new}}(j) p^{\text{new}}(\mathbf{x}^n|j)}{p^{\text{old}}(\mathbf{x}^n) P^{\text{old}}(j|\mathbf{x}^n)} \right\}. \tag{2.91}$$

We wish to minimize E^{new} with respect to the 'new' parameters. If we let Q be the right-hand side in (2.91) then we have $E^{\text{new}} \leq E^{\text{old}} + Q$ and so $E^{\text{old}} + Q$ represents an upper bound on the value of E^{new}. We can therefore seek to minimize this bound with respect to the 'new' values of the parameters, as illustrated in Figure 2.13 (Luttrell, 1994). Minimizing Q will necessarily lead to a decrease in the value of the E^{new} unless E^{new} is already at a local minimum.

If we now drop terms which depend only on the 'old' parameters, we can write the right-hand side of (2.91) in the form

$$\tilde{Q} = -\sum_n \sum_j P^{\text{old}}(j|\mathbf{x}^n) \ln \{ P^{\text{new}}(j) p^{\text{new}}(\mathbf{x}^n|j) \} \tag{2.92}$$

and the smallest value for the upper bound is found by minimizing this quantity. If we consider the specific case of a Gaussian mixture model then we have

$$\tilde{Q} = -\sum_n \sum_j P^{\text{old}}(j|\mathbf{x}^n) \left\{ \ln P^{\text{new}}(j) - d \ln \sigma_j^{\text{new}} - \frac{\|\mathbf{x}^n - \boldsymbol{\mu}_j^{\text{new}}\|^2}{2(\sigma_j^{\text{new}})^2} \right\} + \text{const.} \tag{2.93}$$

We can now minimize this function with respect to the 'new' parameters. For the parameters $\boldsymbol{\mu}_j$ and σ_j this minimization is straightforward. However, for the mixing parameters $P^{\text{new}}(j)$ we must take account of the constraint $\sum_j P^{\text{new}}(j) = 1$. This is easily done by introducing a Lagrange multiplier λ and minimizing the function

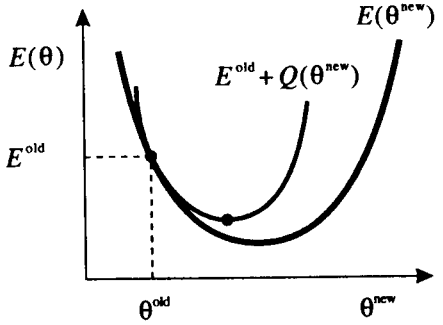

Figure 2.13. Schematic plot of the error function E as a function of the new value θ^{new} of one of the parameters of the mixture model. The curve $E^{\text{old}} + Q(\theta^{\text{new}})$ provides an upper bound on the value of E^{new} and the EM algorithm involves finding the minimum value of this upper bound.

$$\tilde{Q} + \lambda \left(\sum_j P^{\text{new}}(j) - 1 \right). \tag{2.94}$$

Setting the derivatives of (2.94) with respect to $P^{\text{new}}(j)$ to zero we obtain

$$0 = -\sum_n \frac{P^{\text{old}}(j|\mathbf{x}^n)}{P^{\text{new}}(j)} + \lambda. \tag{2.95}$$

The value of λ can be found by multiplying both sides of (2.95) by $P^{\text{new}}(j)$ and summing over j. Using $\sum_j P^{\text{new}}(j) = 1$ and $\sum_j P^{\text{old}}(j|\mathbf{x}^n) = 1$ we obtain $\lambda = N$. We then finally obtain the following update equations for the parameters of the mixture model:

$$\boldsymbol{\mu}_j^{\text{new}} = \frac{\sum_n P^{\text{old}}(j|\mathbf{x}^n) \mathbf{x}^n}{\sum_n P^{\text{old}}(j|\mathbf{x}^n)} \tag{2.96}$$

$$(\sigma_j^{\text{new}})^2 = \frac{1}{d} \frac{\sum_n P^{\text{old}}(j|\mathbf{x}^n) \|\mathbf{x}^n - \boldsymbol{\mu}_j^{\text{new}}\|^2}{\sum_n P^{\text{old}}(j|\mathbf{x}^n)} \tag{2.97}$$

$$P(j)^{\text{new}} = \frac{1}{N} \sum_n P^{\text{old}}(j|\mathbf{x}^n). \tag{2.98}$$

Notice carefully where the 'new' and 'old' parameters appear on the right-hand