A Transparent Interpretation of the EM Algorithm

James M. Coughlan
Smith-Kettlewell Eye Research Institute
2318 Fillmore St.
San Francisco, CA 94115

1 Introduction

The EM algorithm is typically applied to problems with three kinds of variables: an unknown model parameter θ we are trying to estimate, data z which has been observed, and an unobserved variable y, whose value is unknown and should be marginalized over.

The goal is to perform maximum likelihood estimation of θ, i.e. to find the θ which maximizes the log likelihood $\log P(z|\theta)$. The point of this note is to re-express this maximization in a novel form which is equivalent to the interpretation of EM derived by Neal and Hinton [1]. This new formulation makes it obvious that maximizing Neal and Hinton's joint function of θ and a distribution on y is equivalent to maximum likelihood estimation.

The key point is to note that maximizing $\log P(z|\theta)$ over θ is equivalent to maximizing

$$\log P(z|\theta) - D(\tilde{P}(y)||P(y|z,\theta)) \tag{1}$$

jointly over θ and $\tilde{P}(y)$. Here $\tilde{P}(y)$ is any probability distribution on y and $D(p(y)||q(y)) = \sum_y p(y)\log(p(y)/q(y))$ is the Kullback-Leibler divergence between $p(y)$ and $q(y)$. The non-negativity of the Kullback-Leibler divergence, combined with the fact that the divergence is zero only between identical distributions, ensures that the maximum is reached only by setting $\tilde{P}(y)$ equal to the true distribution on y, i.e. $P(y|z,\theta)$.

Equation (1) can be re-written as $H(\tilde{P}) + \sum_y \tilde{P}(y)\log\{P(y|z,\theta)P(z|\theta)\}$, where $H(\tilde{P}) = -\sum_y \tilde{P}(y)\log\tilde{P}(y)$ is the entropy of $\tilde{P}(y)$. This expression is in turn equivalent to

$$H(\tilde{P}) + \sum_y \tilde{P}(y)\log P(y,z|\theta), \tag{2}$$

which is the same as the function $F(\tilde{P},\theta)$ given in Neal and Hinton. This function is maximized iteratively, where each iteration consists of two separate maximations, one over θ and another over $\tilde{P}(y)$.

References

[1] R. Neal and G. Hinton. "A New View of the EM Algorithm that Justifies Incremental and Other Variants." *Biometrika*. 1993.