

# On the Convergence Properties of the EM Algorithm

The paper “On the Convergence Properties of the EM Algorithm” by C. F. Jeff Wu, published in *The Annals of Statistics* in March 1983, delves into a meticulous analysis of the convergence behaviors of the Expectation-Maximization (EM) algorithm. This algorithm is pivotal in the realm of statistical computation, particularly for estimating maximum likelihood parameters within statistical models plagued by incomplete data. The study is bifurcated into examining two principal concerns of convergence: firstly, the capability of the EM algorithm to locate either a local maximum or a stationary point of the likelihood function, and secondly, the convergence trait of the sequence of parameter estimates engendered by the EM algorithm.

Central to the discussion is the EM algorithm’s role in facilitating the estimation process in scenarios where direct maximization of the likelihood function is impractical due to the presence of incomplete data. The algorithm operates through a cyclic process comprising an expectation step (E-step) and a maximization step (M-step). In the E-step, the expected log-likelihood is computed, contingent on the observed data and current estimates of the parameters. This is followed by the M-step, where this expectation is maximized to update the parameter estimates.

Wu critically evaluates the original convergence proof presented by Dempster, Laird, and Rubin (1977), pointing out a significant oversight and proceeding to offer a broader and more rigorous analysis of the algorithm’s convergence characteristics. This exploration is grounded in the principles of optimization theory, considering the EM algorithm as a specialized optimization methodology.

One of the seminal outcomes highlighted in the paper is that under specific conditions, all limit points of the EM sequence emerge as stationary points of the likelihood function. This implies that the sequence of likelihood values converges to a point that, although may not represent the global maximum, is stationary. Mathematically, this can be denoted as

$$L(\theta^{(n)}) \rightarrow L(\theta^*) \text{ as } n \rightarrow \infty,$$

where  $L(\theta)$  is the likelihood function,  $\theta^{(n)}$  is the parameter estimate at the  $n$ th iteration, and  $\theta^*$  is a stationary point.

Additionally, the investigation reveals that if the likelihood function exhibits unimodality and satisfies certain differentiability prerequisites, the EM sequence

converges to the unique maximum likelihood estimate. This finding underscores the practical applicability of the EM algorithm across a wide array of statistical models, particularly those involving incomplete data. In formal terms, if  $L(\theta)$  is unimodal and  $\nabla_{\theta}L(\theta)$  exists and is continuous, then the sequence  $\{\theta^{(n)}\}$  generated by the EM algorithm converges to  $\theta^*$ , the unique maximizer of  $L(\theta)$ .

Wu’s paper culminates in the elucidation of several properties and conditions that assure the EM algorithm’s convergence to a local or global maximum, or to a stationary point. These encompass continuity conditions on the Q-function (integral to the E-step) and the compactness and differentiability of the parameter space. This comprehensive analysis significantly enhances the theoretical framework surrounding the EM algorithm’s convergence properties, offering a robust foundation for its deployment in statistical estimation tasks where incomplete data is a central challenge. Through this work, the reliability and efficiency of the EM algorithm in statistical applications are substantially bolstered, enabling practitioners to leverage it with increased assurance in its convergence behavior.