# Approximate Expectation Maximization

Tom Heskes Onno Zoeter Wim Wiegerinck

SNN, University of Nijmegen, The Netherlands

The full version of this paper appeared in *Advances in Neural Information Processing Systems (NIPS) 16, eds. Thrun, S. and Saul, L. and Schölkopf, B., MIT Press, Cambridge, MA, pp. 353–360, 2004.*

## 1 Expectation Maximization

The EM (expectation-maximization) algorithm [1] is a popular method for maximum likelihood learning in probabilistic models with hidden variables. The E-step boils down to computing probabilities of the hidden variables given the observed variables (evidence) and current set of parameters. The M-step then, given these probabilities, yields a new set of parameters guaranteed to increase the likelihood.

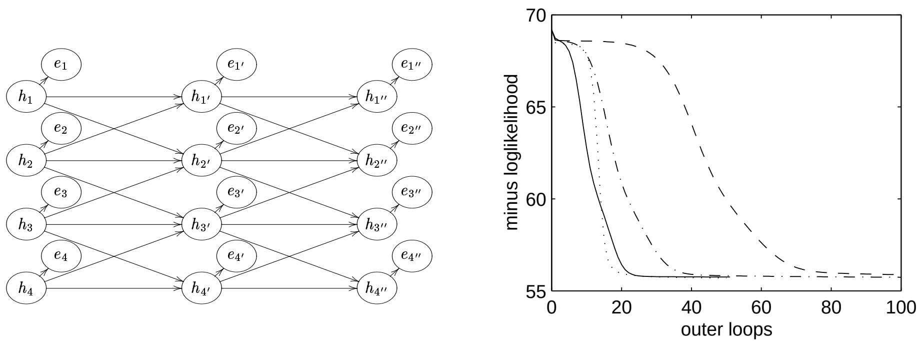
In large probabilistic models, the E-step may become intractable. One then often applies some approximation in the E-step, either through Monte Carlo sampling or with a deterministic variational method. Loopy belief propagation [2] and variants thereof can and have been used for that purpose. A problem, however, is that standard application of these belief propagation algorithms does not always lead to convergence. So-called double-loop algorithms with convergence guarantees are an order of magnitude slower than standard belief propagation.

The goal of this article is to integrate expectation-maximization with belief propagation. This integration directly follows from the free-energy interpretation of EM and belief propagation. Using the Kikuchi approximation instead of the exact free energy, one obtains an approximate EM algorithm that is equivalent to running (some kind of) loopy belief propagation in the E-step combined with an exact M-step.

## 2 From triple-loop to double-loop algorithm

| naive possibly-nonconvergent | triple-loop convergent | double-loop convergent |
|---|---|---|
| { *E-step* ↺??? , *M-step* } ↺ | { { *inner-loop E-step* ↺ , *outer-loop E-step* } ↺ , *M-step* } ↺ | { *inner-loop E-step* ↺ , { *outer-loop E-step* , *M-step* } } ↺ |

The naive approach (on the left, where ↺ stands for "run until convergence") is to just run loopy belief propagation in the E-step, hoping that it converges. With a guaranteed-to-converge double-loop scheme for the E-step, one would get the triple-loop approach in the middle. The crucial observation made in this paper is that for many models the outer-loop E-step can be combined with the M-step, yielding the *convergent double-loop* algorithm on the right. This combination




(a) Coupled hidden Markov model.

(b) Simulation results.

Figure 1: Learning a coupled hidden Markov model. (a) Sketch of architecture for 3 time slices and 4 hidden nodes per time slice. All nodes binary; observed nodes shaded. (b) Minus the loglikelihood as a function of the number of M-steps. Naive algorithm (solid), convergent double-loop algorithm (dashed), optimized convergent algorithm (dash-dotted), same for a Kikuchi approximation (dotted).

makes the double-loop convergent algorithm an order of magnitude faster than the triple-loop convergent algorithm, without giving in on the convergence guarantee.

## 3 Empirical evaluation

One would expect the naive algorithm, when it converges, to be a little faster then the convergent double-loop algorithm. This is verified experimentally when learning a so-called coupled hidden Markov model (see Figure 1(a)) consisting of 20 time slices with 5 hidden nodes per time slice. However, with additional improvements, the convergent algorithm is almost as fast as the naive one (see Figure 1(b)).

The true value of the convergent algorithm becomes clear when the naive algorithm fails. Here we did not manage to get the single-loop version of generalized belief propagation to converge. The convergent algorithm had no problem and yielded the dotted line in Figure 1(b) for the same problem instance.

## References


[1] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[2] K. Murphy, Y. Weiss, and M. Jordan. Loopy belief propagation for approximate inference: An empirical study. In K. Laskey and H. Prade, editors, *Proceedings of the Fifteenth Conference on Uncertainty in Articial Intelligence*, pages 467–475, San Francisco, CA, 1999. Morgan Kaufmann Publishers.