Ming Hsieh Department of Electrical Engineering

USC Viterbi School of Engineering

The Noisy Expectation Maximization Algorithm

Osonde Osoba, Sanya Mitaim, Bart Kosko

Abstract

❑ We present a noise-injected version of the Expectation-Maximization (EM) algorithm: the Noisy Expectation Maximization (NEM) algorithm.

❑ The NEM theorem shows that additive noise speeds up the average convergence of the EM algorithm to a local maximum of the likelihood surface if a positivity condition holds. The NEM algorithm uses the NEM theorem to speed up the convergence of the noiseless EM algorithm.

❑ We have demonstrated these noise benefits on many EM models. We present two models here: the Gaussian mixture model (GMM) and the censored log-convex gamma model.

❑A corollary to the NEM theorem shows that the noise benefit is most pronounced for sparse data.

Background & Theory

❑ Dempster et al.'s EM algorithm is a generalized iterative maximum likelihood method for incomplete or corrupted data. It runs the following two steps on the data samples, y, until the parameter converges:

Algorithm 1 $\hat{\theta}_{EM}$ = EM-Estimate($\vec{y}$)

1: **E-Step:** $Q(\theta|\theta_k) \leftarrow \mathbb{E}_{Z|Y,\theta_k}[\ln f(\vec{y}, z|\theta)]$

2: **M-Step:** $\theta_{k+1} \leftarrow \text{argmax}_{\theta}\{Q(\theta|\theta_k)\}$

❑ We proved the NEM theorem that says that we can improve each iteration of the EM algorithm if we add noise, N, subject to the following positivity condition (NEM condition):

$$\mathbb{E}_{Y,Z,N|\theta^*}\left[\ln \frac{f(Y+N, Z|\theta_k)}{f(Y, Z|\theta_k)}\right] > 0$$

NEM Algorithm

$\mathbf{N_S}$**-Step:** $\vec{n} \leftarrow k^{-\tau} \times \text{NEMNoiseSample}(\vec{y})$

$\mathbf{N_A}$**-Step:** $\vec{y}_\dagger \leftarrow \vec{y} + \vec{n}$

E-Step: $Q(\theta|\theta_k) \leftarrow \mathbb{E}_{Z|Y,\theta_k}[\ln f(\vec{y}_\dagger, z|\theta)]$

M-Step: $\theta_{k+1} \leftarrow \underset{\theta}{\text{argmax}}\{Q(\theta|\theta_k)\}$

❑ The NEM algorithm modifies the EM scheme by introducing a noise-sampling (N_s) and noise-addition (N_A) step

❑ We constrain the noise in the N_s-step subject to the NEM condition. The exact noise constraint depends on the data model.

❑ We also apply a deterministic cooling factor, $k^{-\tau}$, to the noise to eliminate excessive parameter jitter as the algorithm converges.

Noise Benefit in EM Models

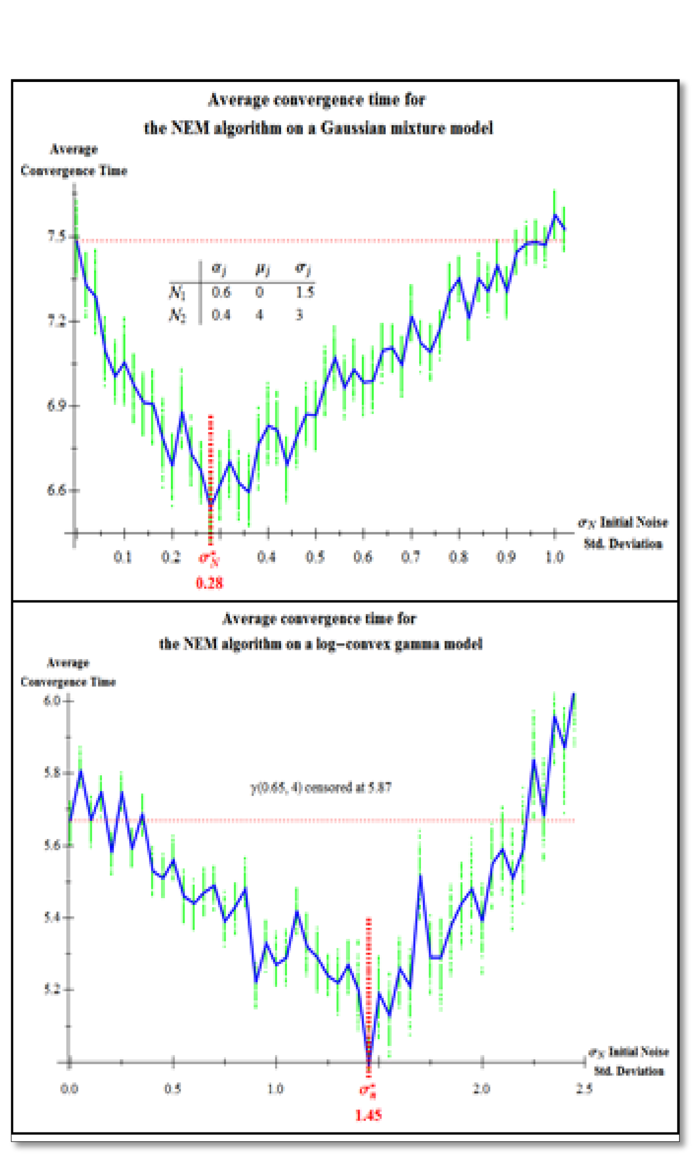
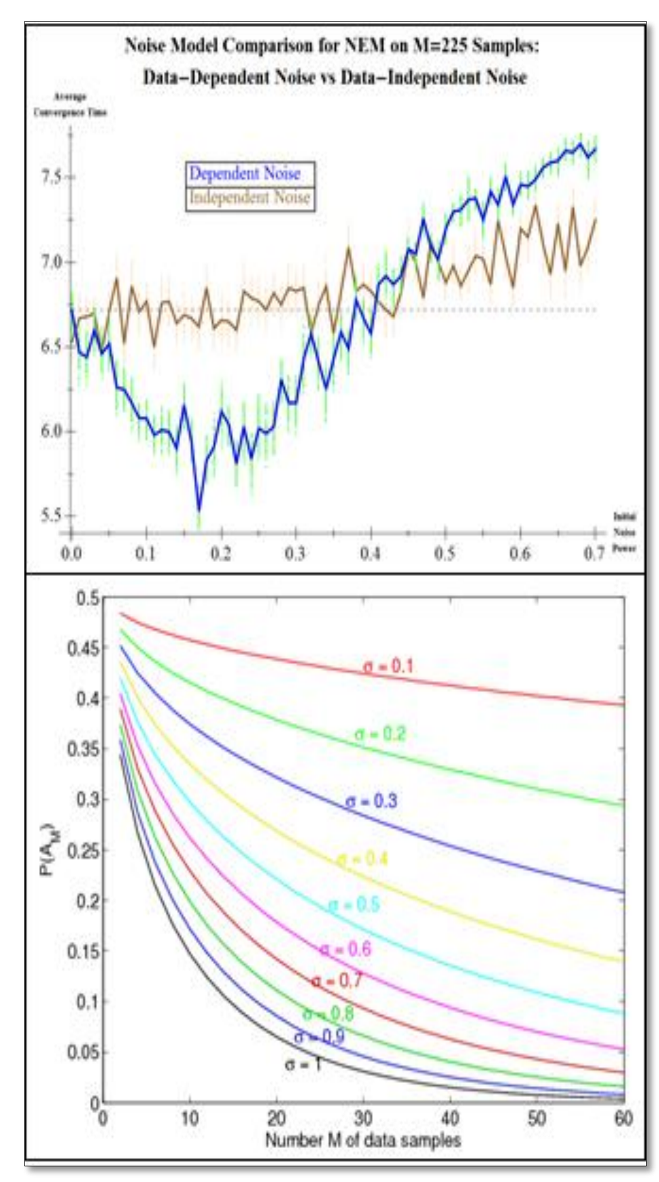


Figure 1: EM noise benefit for a GMM and censored gamma model. Low power initial noise decreases convergence time while higher power initial noise increases it. The average optimal NEM speed-up over the noiseless EM algorithm is about 12% in each case. This NEM additive noise cools at an inverse-square rate. The simulation estimates the scale parameter in the censored-gamma case and the sub-population standard deviations in the GMM case. Each sampled point on the curve is the average convergence time with its associated 95% bootstrap confidence band.

Sparse Data Effects



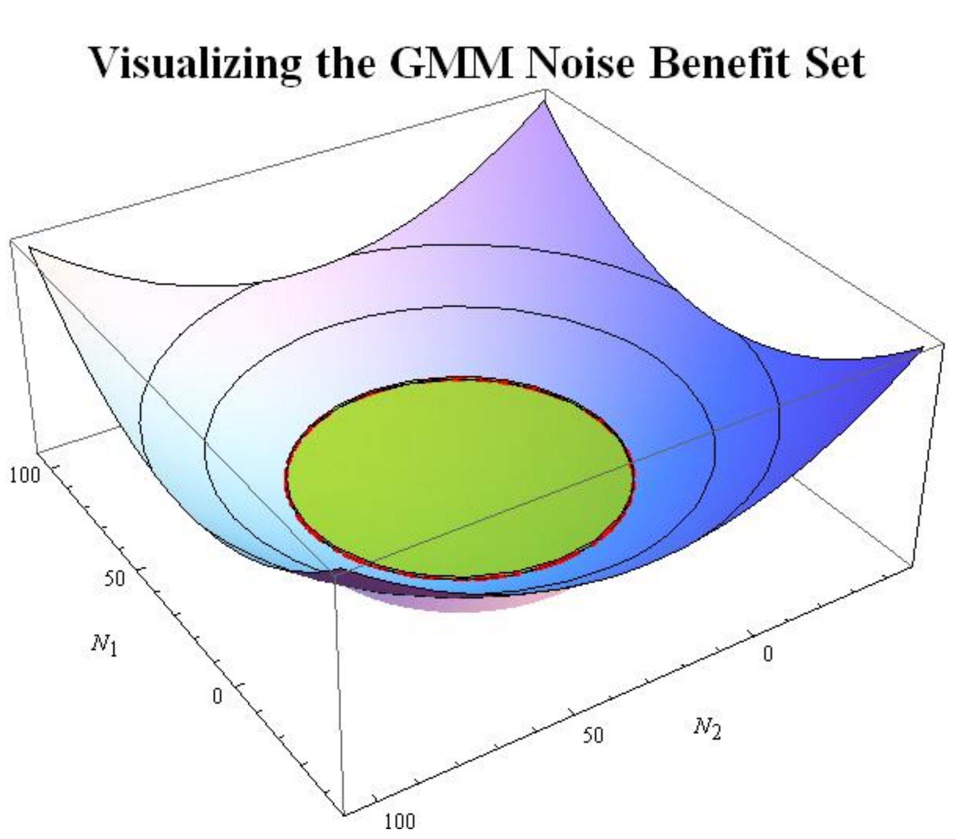
❑ The noise benefit usually disappears if we inject noise without enforcing the NEM condition (top panel).

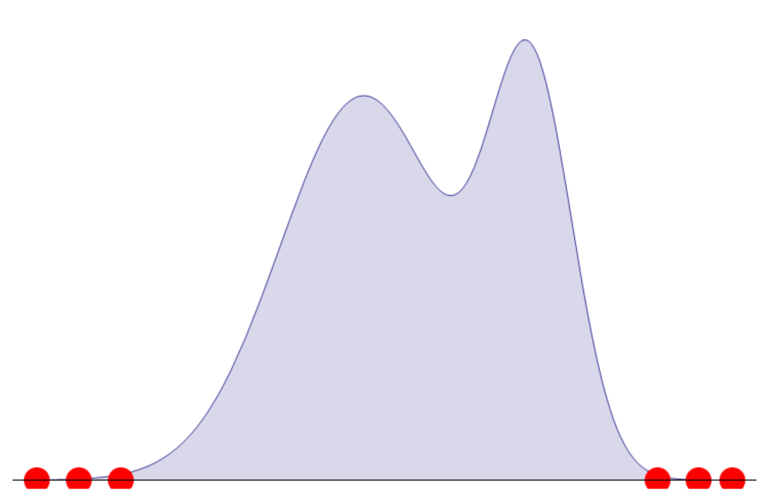
❑ But a NEM theorem corollary states that blind noise addition can be beneficial if the data is sparse (see bottom panel).

Figure 2: The top panel shows the effects of blind noise vs. noise subject to NEM condition: Blind noise gives no benefit while NEM conditioned noise gives a benefit. The bottom panel shows that the probability of satisfying the NEM condition with blind noise samples decreases as sample size increases.

GMM NEM Condition in Detail

The NEM condition for the GMM case specifies the support of the noise pdf. The practical effect of the condition is to make "extreme" data points less "extreme."





The GMM NEM noise pdf support set has the shape of the interior + boundary of a data-dependent hyper-ellipsoid. The figure to the left shows an example for 2-D noise samples for 2-D data.

Conclusion

❑ Careful noise injection can the speed the average convergence time of the EM algorithm. The various sufficient conditions for such a noise benefit involve a direct or average effect where the noise makes the signal data more probable.

❑ Many signal-processing applications apply the EM algorithm to underlying incomplete data models. Such data models include mixture models for data-clustering algorithms (like k-means) and medical image analysis, hidden Markov models (HMM) for automated speech processing and bioinformatics, etc

❑The NEM theorem applies to each of these EM algorithms. Thus these applications can gain speed boosts via specialized NEM algorithms.

❑ Future work should assess noise benefits in other EM applications and develop methods for finding optimal noise levels for NEM algorithms.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," JRSS, Series B, vol. 39, no. 1, pp. 1–38, Jan. 1977

[2] O. Osoba, S. Mitaim, B. Kosko, "Noise Benefits in the Expectation-Maximization Algorithm: NEM Theorems and Models," International Joint Conference on Neural Networks, Jul. 2011

[3] O. Osoba, B. Kosko, "Noise-enhanced Clustering and Competitive Learning Algorithms," Neural Networks, Jan. 2013

[4] O. Osoba, S. Mitaim, B. Kosko, "The Noisy Expectation-Maximization Algorithm," [in review]

Contact: kosko@usc.edu

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering