# A Q-EM BASED SIMULATED ANNEALING ALGORITHM FOR FINITE MIXTURE ESTIMATION

Wenbin Guo and Shuguang Cui

Department of ECE, University of Arizona Tucson, AZ, 85721 Emails: {gwb, cui}@ece.arizona.edu

## ABSTRACT

We develop a *q*-Expectation Maximization (*q*-EM) simulated annealing method for parameter estimation. The *q*-EM algorithm is a one-parameter generalization of normal Expectation Maximization (EM) algorithm based on Tsallis entropy. By incorporating the simulated annealing method, we propose the *q*-Deterministic Annealing Expectation Maximization (*q*-DAEM) algorithm. Given the inherent connection between a physical annealing process and statistical mechanics, we show that the proposed algorithm actually minimizes a counterpart of the free energy in statistical mechanics by controlling an effective temperature. Simulations of mixed Gaussian parameter estimation show that the proposed method is much less initialization-dependent than the standard EM algorithm and converges dramatically faster than the DAEM algorithm.

*Index Terms*— Expectation Maximization, Simulated Annealing, DAEM, Tsallis Entropy, Estimation.

#### 1. INTRODUCTION

Finite mixture model is a flexible and powerful probabilistic tool for statistical learning from incomplete data, with applications such as probability density estimation, clustering, classification, and data summarization [1]. The EM algorithm developed by Dempster [2] is a general method to estimate the parameters of a stochastic model from incomplete data. The fitting of mixture models by maximizing the likelihood is a typical example application of EM algorithms. The standard EM algorithm is highly sensitive to initialization and may converge to a local maximum. A natural solution is to start from multiple random initializations and choose the final estimate with the highest likelihood [3]. Many other adaptations and extensions of the EM approach have been proposed, including the component-wise EM [1] and the split-and-merge EM [4]. Recently, other powerful algorithms such as the genetic algorithm (GA) [5] and the simulated annealing (SA) algorithm [6] are combined with the EM algorithm to combat the issue of local maxima and reduce the sensitivity to initializations

As known, SA is a stochastic technique originated in the study of condensed matter physics. The annealing denotes a physical process where a solid in a heat bath is first heated by increasing the temperature of the heat bath to a maximum value at which all particles in the solid randomly arrange themselves in the liquid phase, and then cooled down by slowly lowering the temperature of the heat bath. In this way, all particles arrange themselves in the low energy ground state of a corresponding lattice. The associated special entropy-changing pattern motivates people to apply simulated annealing method solve statistical optimization problems in [7], where it leads to a greedy search that accepts all changes leading to improvements. However, different from GA, SA allows the acceptance of changes leading to temporary worse solutions with certain probability. By controlling the system parameters such as temperature, SA can find the optimal solution in non-convex cases with high probability, and it has been applied to many problems such as the traveling salesman problem [9] and gene clustering [10].

In addition, a deterministic annealing EM (DAEM) algorithm is proposed in [11], with dramatically improved performance over EM. In [12], a generalized DAEM based on non-extensive statistical mechanics is proposed to further improve the performance. Motivated by the success of Tsallis information theory in the non-extensive statistical physics [13, 14], we have developed the q-EM algorithm in the previous work [15] for joint channel estimation and detection. The q-EM algorithm is a one-parameter generalization of the standard EM algorithms. By controlling the parameter q, we can achieve faster estimation convergence and better detection performance. In this paper, we combine the q-EM algorithm and the simulated annealing method to solve finite mixture estimation problems. We investigate the relationship between q-EM and statistical mechanics, and propose a new algorithm named q-DAEM. Different from the DAEM in [12], we explore the extra freedom offered by the parameter q and apply different constraints. We compare the performance among the standard EM, DAEM, and q-DAEM for finite mixture estimation problems. We show that the proposed q-DAEM has a superior performance over others with fast convergence and less sensitivity to initialization.

The rest of the paper is organized as follows. In Section II, we briefly describe the finite mixture model and the standard EM algorithm. In Section III, we introduce the q-DAEM algorithm based on the maximization of Tsallis entropy. In Section IV, statistical mechanics is connected to the proposed q-DAEM. In Section V, simulation results are shown to compare the standard EM, DAEM, and q-DAEM algorithms. Conclusions are drawn in Section VI.

# 2. THE FINITE MIXTURE MODELS AND EM ALGORITHM

In this section, we briefly introduce the finite mixture model and the standard EM algorithm.

## 2.1. The Finite Mixture Model

Let  $\mathbf{Y} = [Y_1, \dots, Y_d]^T$  be a d-dimensional random variable, with  $y = [y_1, \dots, y_d]^T$  representing a sample drawn from  $\mathbf{Y}$ . The mixture model can be written as

$$p(y|\theta) = \sum_{m=1}^{M} \pi_m p(y|\theta_m) \tag{1}$$

where  $p(y|\theta)$  is the probability density function (PDF) conditional on  $\theta$ ,  $\pi_1, \ldots, \pi_M$  are the mixing probabilities with  $\sum_{m=1}^M \pi_m = 1$ , and  $\theta_m$  is the conditional parameter for the *m*th component, m = $1, \cdots, M$ . Let  $\theta = \{\theta_1, \ldots, \theta_M, \pi_1, \ldots, \pi_M\}$  be the complete set of parameters needed to specify the mixture. Let  $\vec{y} = \{y^{(1)}, \ldots, y^{(n)}\}$ 

This work is supported in part by funds from University of Arizona Foundation and China Scholarship Council.

be the samples of variant **Y**. To understand how each sample  $(y^{(i)})$  is generated, we assume an auxiliary discrete random variable z, which takes M possible values with probability mass  $[p(z = z_1) = \pi_1, \ldots, p(z = z_M) = \pi_M]$ . To generate  $y^{(i)}$ , we first generate a sample of  $z^{(i)}$  based on the above distribution. If  $z^{(i)} = z_m$ , we draw a sample of  $y^{(i)}$  based on  $p(y|\theta_m)$ ,  $m = 1, \ldots, M$ . When y is the received signal in an M-ary digital communication system, z bears physical meaning, and it is actually the M-ary signal constellation point. In such cases, z is called the hidden data. Let us denote the unknown hidden data associated with  $\vec{y}$  as  $\vec{z} = \{z^{(1)}, \ldots, z^{(n)}\}$ . The Maximum log-Likelihood (ML) estimation of  $\theta$  with the observed data  $\vec{y}$  is given by  $\hat{\theta}_{ML} = \operatorname{argmax}_{\theta} \ln p(\vec{y}|\theta)$ .

#### 2.2. The standard EM algorithm

Given the high computational cost of ML algorithms, the standard EM algorithm was proposed in [2] as an efficient iterative solution. The EM algorithm consists of two major steps: Expectation evaluation (E-step), and expectation maximization (M-step). At each iteration, the E-step is with respect to the unknown hidden data, conditional on the current estimated value of the unknown parameters and the observations, as shown below. The M-step then provides a new estimate of the unknown parameters that maximizes the expectation given in the E-step, as shown below. These two steps are iterated until certain convergence criteria are satisfied.

**E-Step**: 
$$Q(\theta, \theta^{(k)}) = E_{\vec{z}}[\ln p(\vec{y}, \vec{z}|\theta)|\vec{y}, \theta^{(k)}].$$

**M-Step**:  $\theta^{(k+1)} = \operatorname{argmax}_{\theta} Q(\theta, \theta^{(k)}).$ 

If the samples of  ${\bf Y}$  are independent, the conditional expectation can be written as

$$Q(\theta, \theta^{(k)}) = \sum_{i=1}^{n} \sum_{m=1}^{M} w_m^{(i)} \ln p(y^{(i)} | z^{(i)}, \theta)$$
(2)

where  $w_m^{(i)} = p(z^{(i)} = z_m | y^{(i)}, \theta^{(k)})$  can be calculated by

$$w_m^{(i)} = \frac{\pi_m^{(k)} p(y^{(i)} | \theta_m^{(k)})}{\sum_{j=1}^M \pi_j^{(k)} p(y^{(i)} | \theta_j^{(k)})}.$$
(3)

Here we see that the EM algorithm is highly dependent on the initialization of  $\theta^{(k)}$ . If the initialization is not close to the global optimum, it may converge to a local one.

#### 3. THE Q-EM BASED SIMULATED ANNEALING

¿From Eq. (3), we see that the conditional probability of  $\vec{z}$  may not be calculated accurately since  $\theta^{(k)}$  itself is an unknown quantity to estimate. Given that we have no prior knowledge about  $p(\vec{z}|\vec{y}, \theta^{(k)})$ , we can apply the *principle of maximum entropy* [16] to have an estimated version, which is deployed in the deterministic annealing EM (DAEM) algorithm [11]. Here, we briefly review the key ideas of DAEM, and then derive the non-extensive DAEM algorithm, which we call *q*-DAEM. Specifically, the conditional probability  $p(\vec{z}|\vec{y}, \theta)$ in DAEM algorithm is determined by solving the following entropy maximization problem:

$$\max \quad H(p) = -\int p(\vec{z}|\vec{y},\theta) \ln p(\vec{z}|\vec{y},\theta) d\vec{z}$$
s. t. 
$$\int p(\vec{z}|\vec{y},\theta) d\vec{z} = 1$$

$$-\int p(\vec{z}|\vec{y},\theta) \ln p(\vec{y},\vec{z}|\theta) d\vec{z} = U$$

$$(4)$$

To solve this problem, we use the Lagrangian multiplier method. In particular, we define

$$J(p) = H(p) + \beta (U + \int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta) \ln p(\vec{\mathbf{y}}, \vec{\mathbf{z}} | \theta) d\vec{\mathbf{z}}) + \lambda (1 - \int p(\vec{\mathbf{z}}, \vec{\mathbf{y}} | \theta) d\vec{\mathbf{z}}),$$

where  $\beta$  and  $\lambda$  are the Lagrange multipliers. By letting  $\frac{\partial J}{\partial p} = 0$ , we have the solution [11]:

$$p(\vec{\mathbf{z}}|\vec{\mathbf{y}},\theta) = \frac{p(\vec{\mathbf{y}},\vec{\mathbf{z}}|\theta)^{\beta}}{\sum_{\vec{\mathbf{z}}} p(\vec{\mathbf{y}},\vec{\mathbf{z}}|\theta)^{\beta}}.$$
 (5)

The DAEM differs from the standard EM algorithm by using the probability  $p(\vec{z}|\vec{y}, \theta)$  calculated in Eq. (5) instead of in Eq. (3). By controlling the  $\beta$  from a small value to one, we obtain the deterministic annealing process of EM [11].

Motivated by the successful applications of non-extensive entropy in long-range interaction physics [13], we have applied the concept to the joint estimation and detection problem in [15]. In this paper, we extend the application to the mixture model estimation problem. Different from [15], there is no feedback from a decoder and the observed data cannot be treated as a hidden Markov process. In addition, as a major modification over DAEM, we determine the conditional probability  $p(\vec{z}|\vec{y}, \theta)$  by maximizing the Tsallis entropy [13] instead of the traditional Shannon entropy:

$$\begin{aligned} \max & H_q(p) = -\frac{1 - \int p(\vec{z}|\vec{y},\theta)^q d\vec{z}}{1 - q} \\ \text{s. t.} & \int p(\vec{z}|\vec{y},\theta) d\vec{z} = 1 \\ & - \int p(\vec{z}|\vec{y},\theta)^q \ln_q(p(\vec{y},\vec{z}|\theta)) d\vec{z} = U_q \end{aligned}$$
(6)

where

$$\begin{aligned} \ln_{q}(x) &\equiv & \frac{x^{1-q}-1}{1-q}, \\ \ln_{q}(\frac{y}{x}) &= & x^{q-1}[\ln_{q}(y) - \ln_{q}(x)], \\ & _{\rightarrow 1}\ln_{q}(x) &= & \ln(x), \end{aligned}$$

and  $U_q$  is a normalization constant. The function  $\ln_q(x)$  is a deformed logarithm function and is monotonic increasing with x. Let us define  $J_q(p)$  as

$$J_{q}(p) = H_{q}(p) + \beta \left( U_{q} + \int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta)^{q} \ln_{q}(p(\vec{\mathbf{y}}, \vec{\mathbf{z}} | \theta) d\vec{\mathbf{z}} \right) + \lambda \left( 1 - \int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta) d\vec{\mathbf{z}} \right),$$
(7)

then we have

 $\lim_{a}$ 

$$\frac{\partial J_q(p)}{\partial p} = \int \left(\frac{q}{1-q}p^{q-1} + \beta q p^{q-1} \ln_q p(\vec{\mathbf{y}}, \vec{\mathbf{z}}|\theta) - \lambda\right) d\vec{\mathbf{z}}.$$

By letting  $\partial J_q(p)/\partial p = 0$ , we have

$$p(\vec{\mathbf{z}}|\vec{\mathbf{y}},\theta) = \frac{\exp_q(\beta \ln_q p(\vec{\mathbf{y}},\vec{\mathbf{z}}|\theta))}{Z}$$
(8)

where

$$\exp_q(x) \equiv (1 + (1 - q)x)^{\frac{1}{1 - q}}_{\dagger},$$
 (9)

$$Z = \int \exp_q(\beta \ln_q p(\vec{\mathbf{y}}, \vec{\mathbf{z}} | \theta)) d\vec{\mathbf{z}} = \left(\frac{(1-q)\lambda}{q}\right)^{\frac{1}{1-q}}, \quad (10)$$

and  $(x)_{\dagger}$  means that  $(x)_{\dagger} = 0$  if x < 0, otherwise  $(x)_{\dagger} = x$ . When  $q, \beta \to 1, Z \to p(\vec{y}|\theta)$ , which is the likelihood function.

After we obtain the estimated  $p(\vec{z}|\vec{y},\theta)$ , we are ready to describe the *q*-DAEM algorithm. Let us denote the *q*-expectation as

$$E_x^{(q)}[g(x)] \equiv \int p(x)^q g(x) dx, \qquad (11)$$

and the algorithm is as follows:

Set β > 1 and q > 1;
 E-Step: Q(θ, θ<sup>(k)</sup>) = E<sup>(q)</sup><sub>z</sub>[ln<sub>q</sub> p(**y**, **z**]θ)|**y**, θ<sup>(k)</sup>];
 M-Step: θ<sup>(k+1)</sup> = argmax<sub>θ</sub>Q(θ, θ<sup>(k)</sup>);
 Decrease q: if q > 1, go back to Step 2;
 If the convergence criteria are not satisfied, decrease β: if β > 1, repeat the above procedures; otherwise stop.

In our experiments, q and  $\beta$  decrease according to  $q(k) = 1 + q(1)e^{-\frac{10k}{L_q}}$  and  $\beta(n+1) = 1 + \beta(n)e^{-\frac{5n}{L_b}}$ , where k and n are the counter index for the inner loop and the outer loop, respectively,  $L_q$  is the maximum allowed number of loops for q, and  $L_b$  is the maximum allowed number of loops for  $\beta$ . We also set q(1) = 10,  $\beta(1) = 5$ ,  $L_q = 20$ , and  $L_b = 10$  for our simulations.

# 4. CONNECTIONS WITH STATISTICAL MECHANICS

In the previous section we proposed the *q*-DAEM algorithm without an intuitive explanation why it works well. In this section, we will show that given the inherent connection between a physical annealing process and statistical mechanics, the proposed algorithm actually minimizes a counterpart of the free energy in statistical mechanics by controlling an effective temperature. It will be proved that minimizing the free energy counterpart implies the maximization of likelihood.

Statistical mechanics is an application of statistics [16], dealing with the motions of a large population of particles or objects that are subjected to external forces. It provides a framework for relating the microscopic properties of individual atoms and molecules to the macroscopic properties of materials that can be observed. This ability to make macroscopic predictions based on microscopic properties is the main asset of statistical mechanics, and the theories are governed by the laws of thermodynamics through entropy. Here, the principles of statistical mechanics are applied to estimation problems, in which we observe the data (macroscopic) that are generated by certain hidden data (microscopic) with a probabilistic model. The parameters to estimate fully describe the statistical behavior.

Taking the deformed logarithm on both sides of Eq. (8), we have

$$\begin{aligned} u_q \, p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta) &= \ln_q \left( \frac{\exp_q(\beta \ln_q p(\vec{\mathbf{y}}, \vec{\mathbf{z}} | \theta))}{Z} \right) \\ &= Z^{q-1} \left( \beta \ln_q p(\vec{\mathbf{y}}, \vec{\mathbf{z}} | \theta) - \ln_q Z \right), \quad (12) \end{aligned}$$

which leads to

lr

$$-\frac{1}{\beta}Z^{1-q}\ln_q p(\vec{\mathbf{z}}|\vec{\mathbf{y}},\theta) = -\ln_q p(\vec{\mathbf{y}},\vec{\mathbf{z}}|\theta) + \frac{1}{\beta}\ln_q Z.$$
(13)

Taking the conditional *q*-expectation over the conditional probability of  $\vec{z}$  on both sides of Eq. (13) and applying Eq. (10), we obtain

$$W(\theta) = U_q - \frac{(1-q)\lambda}{q\beta} H_q(\theta)$$
  
=  $U_q - T_q H_q(\theta)$  (14)

where  $T_q = \frac{(1-q)\lambda}{q\beta}$  is defined as an effective temperature, and

$$W(\theta) \equiv -\frac{1}{\beta} \int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta)^q \ln_q Z d\vec{\mathbf{z}}$$
(15)

$$H_q(\theta) = -\int p(\vec{\mathbf{z}}|\vec{\mathbf{y}},\theta)^q \ln_q p(\vec{\mathbf{z}}|\vec{\mathbf{y}},\theta) d\vec{\mathbf{z}}.$$
 (16)

Here, we see that Eq. (14) has a similar structure to the statistical mechanics modeled by E = U - TS [11], where E is the free energy, U is the total energy, T the temperature, and S is the system entropy. At equilibrium, the thermodynamic system settles into a configuration that minimizes its free energy. With this motivation, we consider the minimization of  $W(\theta)$  by controlling the effective temperature  $T_q$ . To solve this, we derive an iterative algorithm. Let us denote the current estimation as  $\theta^{(k)}$  and take the conditional q-expectation given the observed data y on both sides of Eq. (13), we obtain

$$W(\theta^{(k)}) = U_q(\theta, \theta^{(k)}) - T_q H_q(\theta, \theta^{(k)})$$
(17)

where

$$H_{q}(\theta, \theta^{(k)}) = -\int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta^{(k)})^{q} \ln_{q} p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta) d\vec{\mathbf{z}}$$

$$U_{q}(\theta, \theta^{(k)}) = -\int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta^{(k)})^{q} \ln_{q} p(\vec{\mathbf{y}}, \vec{\mathbf{z}} | \theta) d\vec{\mathbf{z}} \quad (18)$$

$$W(\theta^{(k)}) = -\frac{1}{\beta} \int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta^{(k)})^{q} \ln_{q} Z d\vec{\mathbf{z}}$$

We now show by the following theorem that the proposed q-DAEM algorithm can monotonically drive down  $W(\theta^k)$ , which implies that the likelihood is increased over iterations and reaches maximum if we let  $\beta$  and q approach one in a well-controlled manner.

**Theorem 1** Given  $\theta^{(k)}$  and  $\theta^{(k+1)}$  as a value of  $\theta$  that minimizes  $U_q(\theta, \theta^{(k)})$ , we have  $W(\theta^{(k+1)}) \leq W(\theta^{(k)})$ , where the equality holds if and only if  $U(\theta^{(k+1)}, \theta^{(k)}) = U(\theta^{(k)}, \theta^{(k)})$  and  $p(\vec{z}|\vec{y}, \theta^{(k+1)}) = p(\vec{z}|\vec{y}, \theta^{(k)})$ .

**Proof:** 

$$\begin{split} \Delta W(\theta) &\equiv W(\theta^{(k+1)}) - W(\theta^{(k)}) \\ &= (U_q(\theta^{(k+1)}, \theta^{(k)}) - U_q(\theta^{(k)}, \theta^{(k)})) \\ &+ T_q(-H_q(\theta^{(k+1)}, \theta^{(k)}) + H_q(\theta^{(k)}, \theta^{(k)})) \\ &\leq 0 \end{split}$$

where we used the fact that

$$\begin{aligned} &- H_q(\theta^{(k+1)}, \theta^{(k)}) + H_q(\theta^{(k)}, \theta^{(k)}) \\ &= \int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta^{(k)})^q (\ln_q p(\vec{\mathbf{y}} | \vec{\mathbf{z}}, \theta^{(k+1)}) - \ln_q p(\vec{\mathbf{y}} | \vec{\mathbf{z}}, \theta^{(k)})) d\vec{\mathbf{z}} \\ &= \int p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta^{(k)}) \ln_q \frac{p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta^{(k+1)})}{p(\vec{\mathbf{z}} | \vec{\mathbf{y}}, \theta^{(k)})} d\vec{\mathbf{z}} \\ &\leq 0, \end{aligned}$$

which is based on the property of the *q*-relative entropy proved in [14]. The equality holds if  $p(\vec{z}|\vec{y}, \theta^{(k+1)}) = p(\vec{z}|\vec{y}, \theta^{(k)})$ .

Based on Eq. (10), we see that when both  $\beta$  and q approach one, Z approaches the likelihood  $p(\vec{y}|\theta)$ . Given this fact and the last equation in Eq. (18), we see that the decreasing of  $W(\theta^{(k)})$ is caused by the increasing of likelihood. According to the above theorem, since in our proposed q-DEAM algorithm the condition  $U_q(\theta^{(k+1)}, \theta^{(k)}) \leq U_q(\theta^{(k)}, \theta^{(k)})$  is guaranteed at the M-step, the value of  $W(\theta^{(k)})$  will be decreased while the likelihood is increased over iterations. The convergence speed is controlled by both  $\beta$  and q, which together define  $T_q$  that is an analogy to the temperature in the physical annealing statistical mechanics.

#### 5. SIMULATIONS

In the simulation, the mixture model is given by  $p(y|\theta) = 0.2N(-2, 1) + 0.3N(4, 0.2) + 0.5N(8, 1)$  where  $N(\mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-\mu_i)^2}{2\sigma_i^2}}$ , and the parameters to estimate are  $\theta = \{\pi = [0.2, 0.3, 0.5], \mu = [-2, 4, 8], \sigma^2 = [1, 0.2, 1]\}$ . For each algorithm in comparison, we start from the same initialization and compare certain convergence criteria, which will be defined shortly. Here, we adopt the mean squared error (MSE) between the estimated value and the real value to quantify the estimation process. Let us denote the MSE as  $\varepsilon = E \parallel \hat{\theta} - \theta \parallel^2$ . We assume that the convergence is achieved when  $\varepsilon \leq \delta$ , with  $\delta = 0.1$ . It should be noted that determining convergence in this way may not be practical when the exact value for  $\theta$  is not known. In such case, we may use the mean squared error between the adjacent iterations. The convergence ratio is calculated as the percentage of successful estimations among the total experiments. The convergence speed is defined as the average number of

Table I. Comparison of EM, DAEM, $q$ -DAEM					
Perform	nance	EM	DAEM	q-DAEM	
C	D ··	40.00/	(0 70/	70 10/	

Convergence Ratio	42.3%	69.7%	72.1%
Convergence Speed	53.1	93.5	14.0



Fig. 1. The convergent speed comparison

iterations needed to achieve convergence:  $S_v = \frac{\sum_{i=1}^N C_i}{N}$ , where N is the total number of experiments and  $C_i$  is the actual number of iterations executed to achieve convergence at the *i*th experiment.

For DAEM algorithm, we start the  $\beta$  from 0.1 with the increasing rule as  $\beta = \beta \times 1.2$  [11]. For each experiment, we start from an initialization of  $\sigma^2 = [1, 1, 1]$ ,  $\pi = [1/3, 1/3, 1/3]$ , and  $\mu = [\mu_1, \mu_2, \mu_3]$  that is randomly drawn from a Gaussian distribution of N(1, 3). The total number of experiments is 5000, where 300 samples are drawn from the mixture Gaussian model in each experiment. The maximum number of iterations is limited to 200.

The simulation results are summarized in Table 1, where we see that q-DAEM has the highest convergence ratio and the fastest convergence speed. With DAEM, the convergence speed is a slower than that of the standard EM, since the cooling temperature should be slowly decreased in a normal annealing process. However, with q-DEAM, at each outer loop we fix  $\beta$  and quickly "cool" the process down in the inner loop by controlling  $q \rightarrow 1$ , which leads to a temporary sub-optimal solution. Then we decrease  $\beta$  at an exponential speed to rerun the inner loop, and we finally obtain the global optimum solution with high probability when  $\beta$  is decreased to one. In Fig. 1, we show the convergence performance of the three algorithms for one particular experiment. The convergence of q-DAEM is obviously much faster than others. The corresponding estimated PDF is shown in Fig. 2. In this specific case, the standard EM is actually failed, while the DAEM algorithm and the q-DAEM algorithm are successful.

#### 6. CONCLUSIONS

In this paper, we proposed the q-DAEM algorithm based on the Tsallis entropy and showed its relationship to statistical mechanics. By controlling the system temperature, we obtained an annealing



Fig. 2. The estimated PDF

process to solve the maximum log-likelihood optimization problem in the context of mixed Gaussian parameter estimation. We showed that the proposed algorithm actually minimizes a counterpart of the free energy in statistical mechanics by controlling an effective temperature. By comparing the proposed *q*-DAEM algorithm with the DAEM algorithm and the standard EM algorithm, we showed that the *q*-DAEM algorithm has a much higher convergence speed and successful ratio. **7. REFERENCES** 

- M. A. T. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 3, pp. 381-396, March 2002.
- [2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc. Ser. B*, Vol. 39, No. 1, pp. 1-38, 1977.
- [3] G. Mclachlan and D. Peel, *Finite Mixture Models*. New York: John Wiley & Sons, 2000.
- [4] J. J. Verbeek, N. Vlassis, and B. J. A. Krose, "Efficient Greedy Learning of Gaussian Mixture Models," *Neural Computation*, Vol. 8, No. 2, pp. 469-485, 2003.
- [5] F. Pernkopf and D. Bouchaffra, "Genetic-Based EM Algorithm for Learning Gaussian Mixture Models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, pp. 1344-1348, Aug. 2006.
- [6] E. S. K. See, "A new novel PET/CT reconstruction algorithm by using prior image model with simulated annealing process," in Proceeding of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, December 13-16, Hongkong, ISPACS2005, pp. 661-664, Dec. 2005.
- [7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, Vol. 220, No. 4598, pp. 671-680, 1983.
- [8] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculation by fast computing machines," *J. Chem. Phys.*, Vol. 21, pp. 1087-1092, 1958.
- [9] K. Binder and D. Stauffer, A Simple Introduction to Monte Carlo Simulations and Some Sepcialized Topics, in Applications of the Monte Carlo method in Statistical Physics. Berlin, Germany: Springer-Verlag, pp.1-36, 1985.
- [10] K. Bryan, P. Cunningham, and N. Bolshakova, "Applications of Simulated Annealing to the Biclustering of Gene Expression Data," *IEEE Trans. on Information Technology in Biomedicine*, Vol. 10, No. 3, pp. 519-526, July 2006.
- [11] N. Ueda and R. Nakano, "Deterministic Annealing EM Algorithm," Pergamon, Neural Networks 11, pp.271-282, 1998.
- [12] K. Tabushi and J. Inoue, "Improvement of EM Algorithm by Means of Non-Extensive Statistical Mechanics," *Neural Net*works for Signal Processing XI, 2001. Proceedings of the 2001 IEEE Signal Processing Society Workshop, pp. 133-142, 2001.
- [13] S. Abe and Y. Okamoto (Edit), Nonextensive Statistical Mechanics and Its Applications. Berlin Heidelberg: Springer-Verlag, 2001.
- [14] S. Furuichi, "On Uniqueness Theorems for Tsallis Entropy and Tsallis Relative Entropy," *IEEE Transactions on Information Theory*, Vol. 51, No. 10, pp. 3638-3645, Oct. 2005.
- [15] W. Guo and S. Cui, "Fast Convergence with q-EM based Blind Iterative Detection," to appear at Asilomar Conference on Signals, Systems, and Computers, Oct. 2006.
- [16] E. T. Jaynes, "Information Theory and Statistical Mechanics," *The Physical Review*, Vol. 106, No.4, pp.620-630, May, 1957.