Constraint Optimized Weight Adaptation for Gaussian Mixture Reduction

HD Chen, KC Chang*

Dept. of SEOR
George Mason University
MS 4A6, 4400 University Dr.
Fairfax, VA 22030
*kchang@gmu.edu

Chris Smith

Decisive Analytics Corporation
Technology Innovation Division
1235 South Clark St., #400
Arlington, VA 22202

ABSTRACT

Gaussian mixture reduction is traditionally conducted by recursively selecting two components that appear to be most similar to each other, and merging them. Different definitions on similarity measure have been used in literature. For the case of one-dimensional Gaussian mixtures, K-means algorithms and some variations are recently proposed to cluster Gaussian mixture components in groups, use a center component to represent all in each group, readjust parameters in the center components, and finally perform weight optimization. In this paper, we focus on multi-dimensional Gaussian mixture models. With a variety of reduction algorithms and possible combinations, we developed a hybrid algorithm with constraint optimized weight adaptation to minimize the integrated squared error (ISE). In additions, with computer simulations, we showed that the proposed algorithm provides an efficient and effective Gaussian mixture reduction performance in various random scenarios.

Keywords: Gaussian mixture reduction, Constraint optimization, Integrated squared error, Kullback-Leibler distance.

1. Introduction

In distributed fusion, when a set of mixture distributions are to be fused, the number of resulting basis functions could increase exponentially. Therefore, developing a good management scheme to combine similar terms or prune away insignificant ones will be useful. In this paper, we first examine several existing GMM reduction algorithms and compare their performance. We then develop a new approach by taking advantages of the state-of-the-art algorithms.

2. Fusion with Gaussian Mixture (GMM)

In a mixture model, a probability distribution is represented as a linear combination of basis functions. A Gaussian mixture model (GMM) can be expressed as,

$$f(x)=\sum_{i=1}^{N}\alpha_i N\left(x;\hat{x}_i,P_i\right) \tag{1}$$

where $\sum_{i=1}^{N}\alpha_i=1$. Assuming two GMMs, $f_1(x)=\sum_{i=1}^{N_1}\alpha_{1i}N\left(x;\hat{x}_{1i},P_{1i}\right)$ and $f_2(x)=\sum_{i=1}^{N_2}\alpha_{2i}N\left(x;\hat{x}_{2i},P_{2i}\right)$, are to be fused with a common prior $f_3(x)=\sum_{i=1}^{N_3}\alpha_{3i}N\left(x;\hat{x}_{3i},P_{3i}\right)$ using a standard fusion formula,

$$f_F(x)=\frac{1}{c}\frac{f_1(x)f_2(x)}{f_3(x)} \tag{2}$$

where $c=\int\frac{f_1(x)f_2(x)}{f_3(x)}dx$ is the normalization constant. With GMM, from (1) and (2), we have,

$$f_F(x)=\frac{1}{c}\frac{\sum_{i=1}^{N_1}\alpha_{1i}N\left(x;\hat{x}_{1i},P_{1i}\right)\sum_{j=1}^{N_2}\alpha_{2j}N\left(x;\hat{x}_{2j},P_{2j}\right)}{p_3(x)}=\frac{1}{c}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\alpha_{1i}\alpha_{2j}\left(\frac{N\left(x;\hat{x}_{1i},P_{1i}\right)N\left(x;\hat{x}_{2j},P_{2j}\right)}{p_3(x)}\right) \tag{3}$$

where

$$c=\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\alpha_{1i}\alpha_{2j}\int\frac{N\left(x;\hat{x}_{1i},P_{1i}\right)N\left(x;\hat{x}_{2j},P_{2j}\right)}{f_3(x)}dx\equiv\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\alpha_{1i}\alpha_{2j}c_{ij} \tag{4}$$

is the normalization constant. In general, the integration in Equation (4) can not be obtained in closed-form. To avoid the potential complexity using numerical integration, we proposed to approximate the denominator, $p_3(x)$, with a single Gaussian [1]. Namely,

$$f_3(x)=\sum_{i=1}^{N_3}\alpha_{3i}N\left(x;\hat{x}_{3i},P_{3i}\right)\approx N\left(x;\hat{x}_3,P_3\right) \tag{5}$$

where $\hat{x}_3=\sum_{i=1}^{N_3}\alpha_{3i}\hat{x}_{3i}$ and $P_3=\sum_{i=1}^{N_3}\alpha_{3i}\left[P_{3i}+\left(\hat{x}_{3i}-\hat{x}_3\right)\left(\hat{x}_{3i}-\hat{x}_3\right)'\right]$.

With this approximation, the integration can be carried out analytically. Specifically,

$$c_{ij} \approx \int \frac{N(x;\hat{x}_{1i},P_{1i})\,N(x;\hat{x}_{2j},P_{2j})}{N(x;\hat{x}_3,P_3)}dx = \sqrt{\frac{|P_3||P_{ij}|}{|P_1||P_2|}}\exp\left(-\frac{1}{2}\right)\left[\hat{x}_{1i}'P_{1i}^{-1}\hat{x}_{1i}+\hat{x}_{2j}'P_{2j}^{-1}\hat{x}_{2j}-\hat{x}_3'P_3^{-1}\hat{x}_3-\hat{x}_{ij}'P_{ij}^{-1}\hat{x}_{ij}\right]$$
$$= \sqrt{\frac{|P_3||P_{ij}|}{|P_1||P_2|}}\exp\left(-\frac{1}{2}\right)\left[\left(\hat{x}_{1i}-\hat{x}_{ij}\right)'P_{1i}^{-1}\left(\hat{x}_{1i}-\hat{x}_{ij}\right)+\left(\hat{x}_{2j}-\hat{x}_{ij}\right)'P_{2j}^{-1}\left(\hat{x}_{2j}-\hat{x}_{ij}\right)-\left(\hat{x}_3-\hat{x}_{ij}\right)'P_3^{-1}\left(\hat{x}_3-\hat{x}_{ij}\right)\right] \quad (6)$$

where $P_{ij}=\left[P_{1i}^{-1}+P_{2j}^{-1}-P_3^{-1}\right]^{-1}$ and $\hat{x}_{ij}=P_{ij}\left[P_{1i}^{-1}\hat{x}_{1i}+P_{2j}^{-1}\hat{x}_{2j}-P_3^{-1}\hat{x}_3\right]^{-1}$.

From (6), one can obtain the normalization constant (4) analytically. Similarly, each term in (3) can also be obtained by the following approximation,

$$\frac{N(x;\hat{x}_{1i},P_{1i})\,N(x;\hat{x}_{2j},P_{2j})}{f_3(x)} \approx \frac{N(x;\hat{x}_{1i},P_{1i})\,N(x;\hat{x}_{2j},P_{2j})}{N(x;\hat{x}_3,P_3)} = c_{ij}\cdot N(x;\hat{x}_{ij},P_{ij}) \quad (7)$$

With that, Equation (3) can be rewritten as,

$$f_F(x) \approx \frac{1}{c}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\alpha_{1i}\alpha_{2j}c_{ij}\cdot N(x;\hat{x}_{ij},P_{ij}) \quad (8)$$

However, as one can see from (8), the fused pdf will have exponentially growing number of basis functions as more pdfs are fused. One way to avoid this computational problem is to combine similar terms or prune away low likely terms.

3. Gaussian Mixture Reduction

Given a Gaussian mixture model (GMM) distribution with *N* components, we wish to approximate it by a Gaussian mixture with *M* components, where $M < N$. Traditionally, a mixture reduction algorithm is recursively conducted such that the number of components is reduced by repeatedly choosing the two components that appear to be most similar to each other and merging them.

West [2] proposed to collapses mixture components by simply replacing nearest neighboring components with a single average component. The basic routine proceeds as follows: First, locate the component with smallest weight. Then find another component, which is the nearest neighbor of the selected one. Finally, merge the two components such that the resulting component is the weighted average of the two. The procedure is repeated until the desirable reduction of components is achieved.

In the literature, various definitions of similarity measures were employed in different algorithms. Williams [3] used integrated squared error (ISE) as the similarity measure. Runnalls [4] used the Kullback-Leibler (KL) discrimination measure. In this report, we will first evaluate existed algorithms, and then propose some new algorithms, including some enhancements/combinations on these existed algorithms.

3.1. Integrated Squared Error

One of the most popular measures to compare the difference between two GMMs is the so called "integrated squared error" (ISE) [2] due to its simple closed forms expressions.

Assumed that the density function of a GMM with N components given in (1) can be reduced to

$$\tilde{f}(x)=\sum_{i=1}^{M}\tilde{\alpha}_i N\left(x;\tilde{x}_i,\tilde{P}_i\right) \tag{9}$$

where M < N. The ISE distance to measure the similarity between two arbitrary Gaussian mixtures is defined as,

$$\begin{aligned} D_{ISE} &= \int_{\Omega}\left[f(\mathbf{x})-\tilde{f}(\mathbf{x})\right]^2 d\mathbf{x} \\ &= \int_{\Omega} f(\mathbf{x})^2 d\mathbf{x}+\int_{\Omega}\tilde{f}(\mathbf{x})^2 d\mathbf{x}-2\int_{\Omega} f(\mathbf{x})\tilde{f}(\mathbf{x})d\mathbf{x} \end{aligned} \tag{10}$$

All three terms in (10) can be derived in closed forms, namely

$$\begin{aligned} \int_{\Omega} f(\mathbf{x})^2 d\mathbf{x} &= \sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j\int_{\Omega}\mathbb{N}(\mathbf{x};\mathbf{u}_i,\mathbf{P}_i)\mathbb{N}\left(\mathbf{x};\mathbf{u}_j,\mathbf{P}_j\right)d\mathbf{x} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j\mathbb{N}\left(\mathbf{u}_i;\mathbf{u}_j,\mathbf{P}_i+\mathbf{P}_j\right) \end{aligned} \tag{11}$$

$$\begin{aligned} \int_{\Omega} \tilde{f}(\mathbf{x})^2 d\mathbf{x} &= \sum_{i=1}^{L}\sum_{j=1}^{L}\tilde{\alpha}_i\tilde{\alpha}_j\int_{\Omega}\mathbb{N}(\mathbf{x};\tilde{\mathbf{u}}_i,\tilde{\mathbf{P}}_i)\mathbb{N}\left(\mathbf{x};\tilde{\mathbf{u}}_j,\tilde{\mathbf{P}}_j\right)d\mathbf{x} \\ &= \sum_{i=1}^{L}\sum_{j=1}^{L}\tilde{\alpha}_i\tilde{\alpha}_j\mathbb{N}\left(\tilde{\mathbf{u}}_i;\tilde{\mathbf{u}}_j,\tilde{\mathbf{P}}_i+\tilde{\mathbf{P}}_j\right) \end{aligned} \tag{12}$$

$$\begin{aligned} \int_{\Omega} f(\mathbf{x})\tilde{f}(\mathbf{x})d\mathbf{x} &= \sum_{i=1}^{M}\sum_{j=1}^{L}\alpha_i\tilde{\alpha}_j\int_{\Omega}\mathbb{N}(\mathbf{x};\mathbf{u}_i,\mathbf{P}_i)\mathbb{N}\left(\mathbf{x};\tilde{\mathbf{u}}_j,\tilde{\mathbf{P}}_j\right)d\mathbf{x} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{L}\alpha_i\tilde{\alpha}_j\mathbb{N}\left(\mathbf{u}_i;\tilde{\mathbf{u}}_j,\mathbf{P}_i+\tilde{\mathbf{P}}_j\right) \end{aligned} \tag{13}$$

3.2. Proposed Algorithm

The proposed algorithm is based on a combination of the WEST algorithm and a constraint optimized weight adaptation (COWA) algorithm. Specifically, the algorithm consists of the following steps:

(1) Apply enhanced WEST algorithm

In the enhanced WEST algorithm, a normalized (by the determinant of the covariance) weight of each component is obtained such that

$$w_i^{'} = \frac{w_i}{\det(\mathbf{P}_i)} \quad (14)$$

These normalized weights are used to order the GMM components such that the component with the smallest weight is selected and merged with another one which is closest to it in the ISE sense. Specifically, the two components are merged based on the following equations,

$$w_{ij} = w_i + w_j \quad (15)$$

$$\mathbf{u}_{ij} = \lambda_i \mathbf{u}_i + \lambda_j \mathbf{u}_j \quad (16)$$

$$\mathbf{P}_{ij} = \lambda_i \mathbf{P}_i + \lambda_j \mathbf{P}_j + \lambda_i \lambda_j (\mathbf{u}_i - \mathbf{u}_j)(\mathbf{u}_i - \mathbf{u}_j)^{\mathrm{T}} \quad (17)$$

where $\lambda_i = w_i / w_{ij}$ and $\lambda_j = w_j / w_{ij}$.

(2) Apply constraint optimized weight adaptation

After each reduction step, with the reduced GMM, apply the constraint optimized adaptation algorithm (COWA) to adjust the GMM component weights such the ISE distance between the reduced GMM and the original GMM is minimized. The details of the adaptation algorithm are presented in the next section.

(3) Repeat the above steps until either the number of components reaches the goal or the ISE distance is above the pre-determined threshold.

3.3. Constraint Optimized Weight Adaptation

Suppose that we have initially located a Gaussian mixture of K components to approximate an original Gaussian mixture of N components, where K<N. We now use constraint optimization method to adapt the K-component weights to minimize the ISE from the original GMM. The general model can be expressed as

$$\min \int_{\Omega} \left[\sum_{i=1}^{N} \alpha_i \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i, \mathbf{P}_i\right) - \sum_{j=1}^{K} \beta_j \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i^{'}, \mathbf{P}_i^{'}\right) \right]^2 d\mathbf{x}$$
$$s.t. \quad \sum_{j=1}^{K} \beta_j = 1 \tag{9}$$

Where $\mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i, \mathbf{P}_i\right)$ denotes the i-th component of multivariate Gaussian density with mean, $\mathbf{u}_i$, and covariance, $\mathbf{P}_i$. The weights satisfy $\sum_{i=1}^{N} \alpha_i = 1$. With any GMM reducing process of changing N into K components (K<N), the reduced GMM may be

$$\tilde{f}\left(\mathbf{x}\right) = \sum_{i=1}^{K} \beta_i \cdot \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i^{'}, \mathbf{P}_i^{'}\right) \tag{10}$$

Where $\sum_{i=1}^{L} \beta_i = 1$. A distance measure commonly used is the ISE, which can be derived in closed forms for the similarity measure between two arbitrary Gaussian mixtures. Our objective is to find the best weights of $\{\beta_i\}$ to minimize ISE. Using the Lagrange formula, we have

$$\left[\sum_{i=1}^{N} \alpha_i \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i, \mathbf{P}_i\right) - \sum_{j=1}^{K} \beta_j \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i^{'}, \mathbf{P}_i^{'}\right) \right]^2 + \lambda \left(\sum_{j=1}^{K} \beta_j - 1 \right) \tag{11}$$

Compute the first derivative of (11) with respective to $\{\beta_i\}$ and equal to zero. We will have the optimal solution in close forms, namely

$$\mathbf{b}^{*} = \mathbf{H}^{-1}\mathbf{a} - \mathbf{H}^{-1}\mathbf{c}\left(\mathbf{c}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{a} - 1\right)\left(\mathbf{c}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{c}\right)^{-1} \tag{12}$$

Where

$$\mathbf{c}^{\mathrm{T}} \equiv \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \tag{13}$$

$$\mathbf{b}^{\mathrm{T}} \equiv \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_k \end{bmatrix} \tag{14}$$

$$\mathbf{a} \equiv \begin{bmatrix} \sum_{i=1}^{N} \alpha_i \int \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i, \mathbf{P}_i\right) \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_1^{'}, \mathbf{P}_1^{'}\right) d\mathbf{x} \\ \vdots \\ \sum_{i=1}^{N} \alpha_i \int \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_i, \mathbf{P}_i\right) \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_k^{'}, \mathbf{P}_k^{'}\right) d\mathbf{x} \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{i=1}^{N} \alpha_i \mathbb{N}\left(\mathbf{u}_1^{'} \mid \mathbf{u}_i, \mathbf{P}_1^{'} + \mathbf{P}_i\right) \\ \vdots \\ \sum_{i=1}^{N} \alpha_i \mathbb{N}\left(\mathbf{u}_k^{'} \mid \mathbf{u}_i, \mathbf{P}_k^{'} + \mathbf{P}_i\right) \end{bmatrix} \tag{15}$$

and

$$\mathbf{H} \equiv \begin{bmatrix} \int \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_1^{'}, \mathbf{P}_1^{'}\right)\mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_1^{'}, \mathbf{P}_1^{'}\right)d\mathbf{x} & \cdots & \int \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_1^{'}, \mathbf{P}_1^{'}\right)\mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_k^{'}, \mathbf{P}_k^{'}\right)d\mathbf{x} \\ \vdots & \ddots & \vdots \\ \int \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_k^{'}, \mathbf{P}_k^{'}\right)\mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_1^{'}, \mathbf{P}_1^{'}\right)d\mathbf{x} & \cdots & \int \mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_k^{'}, \mathbf{P}_k^{'}\right)\mathbb{N}\left(\mathbf{x} \mid \mathbf{u}_k^{'}, \mathbf{P}_k^{'}\right)d\mathbf{x} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbb{N}\left(\mathbf{u}_1^{'} \mid \mathbf{u}_1^{'}, 2\mathbf{P}_1^{'}\right) & \cdots & \mathbb{N}\left(\mathbf{u}_1^{'} \mid \mathbf{u}_k^{'}, \mathbf{P}_1^{'} + \mathbf{P}_k^{'}\right) \\ \vdots & \ddots & \vdots \\ \mathbb{N}\left(\mathbf{u}_k^{'} \mid \mathbf{u}_1^{'}, \mathbf{P}_1^{'} + \mathbf{P}_k^{'}\right) & \cdots & \mathbb{N}\left(\mathbf{u}_k^{'} \mid \mathbf{u}_k^{'}, 2\mathbf{P}_k^{'}\right) \end{bmatrix} \quad (16)$$

4. Preliminary Results

We tested a 2-dimension GMM with 10 components. The prior probability is uniformly distributed. Mean and covariance of each component are randomly and independently generated. We compared three GMM reduction algorithms: enhanced-WEST, KL-based, and AW. 10,000 simulation trails were conducted for performance evaluations. Figure 1 shows the simulation results as components are reduced one-by-one. It is clear that the AW algorithm performs the best in this experiment.

5. Summary

We have conducted some preliminary test. The results so far are very promising. We intend to develop extensive test with a set of representative random examples.

REFERENCES

[2] W. West, “Approximating Posterior Distributions by Mixture”, Journal of the Royal Statistical Society, Series B (Methodological), Vol. 55, No. 2, pp. 409-422, 1993.

[3] J. L. Williams and P. S. Maybeck, “Cost-Function-Based Hypothesis Control Techniques for Multiple Hypothesis Tracking”, in proceedings of the sixth international conference on Information Fusion, Vol.2, pp.1047-1054, 2003.

[4] A. R. Runnalls, “Kullback-Leibler Approach to Gaussian Mixture Reduction”, IEEE Transactions on Aerospace and Electronic Systems, Vol.43, NO.3, July 2007.

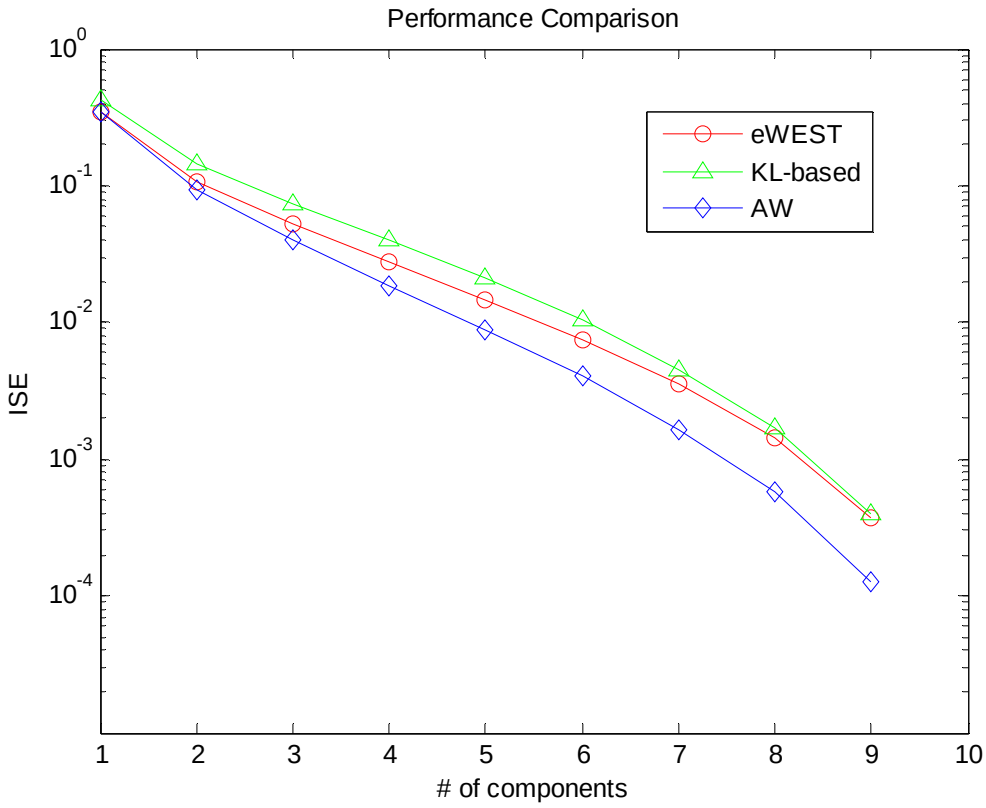

Figure 1. Comparing RSS Grid-based Fusion with GMM Fusion