A Reduced Gaussian Mixture Representation Based On Sparse Modeling

Hongyan Zhu  Chongzhao Han  Yan Lin
Automation department
School of Electronic & Information Engineering, Xi’an Jiaotong University
Xi’an, P. R. China
hyzhu2011@gmail.com

Abstract—Gaussian mixture is an important probability representation for the system state in many applications, especially used to model a complex density. The common drawback of Gaussian mixture representation is that the component number may grow rapidly in case with the recursive processing of Gaussian mixture. Therefore, it is extremely important to develop an efficient Gaussian Mixture Reduction (GMR) strategy to lower the increasing number of mixture components. A GMR approach based on sparse modeling is proposed in this paper. Firstly, we seek to construct a Gaussian base set by merging partial components of the original mixture, from which the components of the reduced mixture will be selected. Secondly, we aim to select a given number of components from the given Gaussian base to form the reduced mixture. By means of the idea of sparse representation, we turn the above selecting problem into a sparse modeling problem successfully in the Framework of $L_{1/2}$ regularization. Finally, we adopt the iterative half thresholding algorithm to acquire the $L_{1/2}$ regularization solution for GMR. Simulation results demonstrated the efficiency of the proposed approach.

Keywords: GMR, $L_{1/2}$ regularization, sparse modeling.

I. INTRODUCTION

Gaussian mixture is an important probability representation for the system state in many applications such as data fusion, pattern recognition, supervised learning of multimedia data and integrated navigation, etc. By using a small number of Gaussian components, one can model distributions that are far from Gaussian. For example, in some target tracking applications with measurement origin uncertainty, the target state is represented as a multivariate Gaussian mixture resulting from the common used data association schemes.

The drawback of such a Gaussian mixture representation is that there is a great tendency for the component number to grow rapidly and without an upper bound in cases with the recursive processing of Gaussian mixture. Therefore, it is necessary to develop an efficient GMR approach with minimal loss of fidelity, so as to meet the challenges of both the computation and storage capability arisen from the rapid increasing of component number. Typically, this can be achieved either by eliminating components with low probability, and/or by merging partial components according to some similarity measure.

For most GMR algorithms, the main focuses involve identifying those components being similar enough and merging them. The pioneering work for GMR was done in [1], which presented a similarity measure based on the statistical analysis of variance and sought to minimize the increase in ‘within-component’ variance resulting from merging two chosen components. The component number is reduced by repeatedly choosing two similar components and merging them [1-2]. Another easily calculated measure of similarity and its extension to higher dimensions between pairs of densities is made by a square-integral function [3-4], which served to find the most similar pair of components and collapse them without comparing all pairs. Another criterion called ISD (integral squared difference, also named as ISE-integral squared error) is adopted in [5-6], which measured the overall error between two Gaussian mixtures instead of the distance between two Gaussian components to achieve a better global performance. The algorithm uses a gradient-based optimization to seek a reduced mixture with a lower dissimilarity to original mixture. A detailed
comparison of various measures of similarity is given in [7].

A K-L divergence-based distance to measure the similarity between two Gaussian mixtures was presented in [8]. Since there appears to be no close-formed expression, the paper put an upper bound on the K-L divergence of two mixtures. In [9], a Gaussian mixture reduction approach by k-means clustering (GMRC) was given, which uses K-L divergence to evaluate the distance between the cluster centers and the merged states, and adopts an iterative optimization process to minimize an upper bound on the K-L divergence between the full mixture and the reduced mixture. In [10], a constraint optimized weight adaptation (COWA) procedure was carried out to refine the weights by minimizing the ISE, if given an initial approximate mixture with desired component number. Instead of reducing the original Gaussian mixture directly, an approximate mixture was produced which allowed to create a mixture from one component and then sequentially add new components each time [11]. A general look at GMR is given in [12]. And also, a brute-force clustering and reduction algorithm was proposed to overcome the sensitivity of initial value existing in original GMRC algorithm [9].

The sparse modeling problems have attracted a great deal of attention in recent years, which aims to find a sparse representation when given some full one. Typically, the sparse modeling problems involve variable selection [13], visual coding [14-15], and compressed sensing [16-17]), etc. The special importance of $L_{1/2}$ regularization [18-20] has been recognized in recent studies on sparse modeling (particularly, on compressed sensing). The $L_{1/2}$ regularization, however, leads to a non-convex, non-smooth and non-Lipschitz optimization problem that is difficult to solve fastly and efficiently. In [19], an iterative half thresholding algorithm for fast solution of $L_{1/2}$ regularization problem is proposed, which corresponds to the well-known iterative soft thresholding algorithm for $L_1$ regularization [21-23], and the iterative hard thresholding algorithm for $L_0$ regularization.

In this paper, we have tried to tackle the GMR problem by means of the idea of sparse modeling. Firstly, a Gaussian base set is needed from which the components of the reduced mixture can be selected. We construct the base by merging partial components of the original mixture by the rule of keeping the 1st and 2nd statistical moments. The key step is to develop a distance measure between two Gaussian mixtures and determine which components will be merged according to this measure. And then, we are in the position to select the components from the base set and assign appropriate weights to form the reduced mixture. By means of the idea of sparse representation, we turn the above selecting problem into a sparse modeling problem successfully in the Framework of $L_{1/2}$ regularization. Finally, we adopt the iterative half thresholding algorithm to acquire the $L_{1/2}$ regularization Solution for GMR.

The remainder of the paper is organized as follows. Section II formulates the original GMR problem and $L_{1/2}$ regularization theory. Section III shows how to construct the Gaussian base set, and how to relax the original GMR problem to be a sparse modeling problem in the framework of $L_{1/2}$ regularization. The performance comparison of the proposed approach and competitive algorithms are given in section IV. We conclude in Section V.

## II PROBLEM FORMULATION

### A. Original GMR Problem

Given a Gaussian mixture density

$$f(x \mid \Omega_N) = \sum_{i=1}^{N} w_i N_i(x)$$

We denote the original Gaussian mixture by $\Omega_N$, that is

$$\Omega_N \triangleq \{w_i, \mu_i, P_i\}_{i=1}^{N}$$

Each component $i \ (i = 1, \ldots, N)$ of the Gaussian mixture is of the following density
The mixture weight \( w = [w_1, ..., w_N] \) satisfies
\[
\sum_{i=1}^{N} w_i = 1, 0 \leq w_i \leq 1
\] (4)

The goal of GMR is to seek a reduced mixture \( N(x | \Omega_K) \) to approximate the original mixture \( N(x | \Omega_N) \) with minimal loss of fidelity.

The density \( f(x | \Omega_K) \) of the reduced mixture is expressed by
\[
f(x | \Omega_K) = \sum_{j=1}^{K} \tilde{w}_j \tilde{N}_j(x)
\] (5)

where
\[
\Omega_K \triangleq \{\tilde{w}_j, \tilde{\mu}_j, \tilde{\Sigma}_j\}_{j=1}^{K}
\] (6)

and
\[
\tilde{N}_j(x) = N(x; \tilde{\mu}_j, \tilde{\Sigma}_j) = \frac{1}{(2\pi)^{n/2} |\tilde{\Sigma}_j|^{1/2}} e^{-\frac{1}{2} (x-\tilde{\mu}_j)^T \tilde{\Sigma}_j^{-1} (x-\tilde{\mu}_j)}
\] (7)

where \( n \) means the dimension of the state vector \( x \).

We choose ISD (integral squared difference) to measure the similarity of two Gaussian mixtures. The big advantage of the criterion is that the distance between two Gaussian mixtures can be expressed in closed form without requiring expensive numerical integration. In this way, the optimization problem can be described as
\[
\min_{\theta} \int_x \left( \sum_{i=1}^{N} w_i N_i(x) - \sum_{j=1}^{K} \tilde{w}_j \tilde{N}_j(x) \right)^2 dx
\] (8)

subject to
\[
\sum_{j=1}^{K} \tilde{w}_j = 1, 0 \leq \tilde{w}_j \leq 1
\] (9)

where \( \theta \triangleq \{\tilde{w}_j, \tilde{\mu}_j, \tilde{\Sigma}_j\}_{j=1}^{K} \) are the parameters to estimate.

**B. \( L_{1/2} \) Regularization**

In this subsection, we review \( L_{1/2} \) regularization theory for sparse modeling, and seek to explore the inner relationship between GMR problem and \( L_{1/2} \) regularization theory. For more details about \( L_{1/2} \) regularization theory, we refer the reader to [18-20].

Given a measurement matrix \( A \) and measurement vector \( y \) with the following linear mapping relationship
\[
y = Ax + \epsilon
\] (10)

where \( \epsilon \) is the observation noise. The sparse modeling problem aims to find the sparse solution of a representation or an equation. That is, to recover \( x \) from the measurement \( y \), such that \( x \) is of the sparest structure (that means \( x \) has the fewest nonzero components).

The problem can be modeled by
\[
\min_{x} \left\{ \|y - Ax\|^2 + \lambda \|x\|_0 \right\}
\] (11)

where \( \|x\|_0 \) denotes \( L_0 \) norm of \( x \) which is the number of nonzero components of \( x \), \( \|\cdot\| \) means Euclidean norm, and \( \lambda \) is a regularization parameter.

Formula (11) is called \( L_0 \) regularization problem. We can consider it as a penalized least square with \( L_0 \) norm \( \|x\|_0 \), in which the parameter \( \lambda \) functions as balancing two objective terms.

The \( L_0 \) regularization is regarded as an idea model for the sparse modeling problems. However, the complexity of the model is proportional to the number of variables which is computationally intractable in general. As an alternative approach, the \( L_1 \) regularization [21-23] gets its popularity since it provides the best convex approximation to the \( L_0 \) regularization and can be solved efficiently. Nevertheless, the \( L_1 \) regularization fails to yield the sparsest solution in most application cases. In particular, it cannot handle the collinearity and may result in an inconsistent selection when it is applied to a
variable selection problem. The special importance of \( L_{1/2} \) regularization has been recognized in recent studies on sparse modeling. That is,

\[
\min_{x \in \mathbb{R}^n} \left\{ \| y - Ax \|^2 + \lambda \| x \|_{1/2}^{1/2} \right\}
\]

(12)

where \( \| x \|_q (0 < q < 1) \) is the \( L_q \) quasi-norm in \( \mathbb{R}^N \) defined by

\[
\| x \|_q = \left( \sum_{i=1}^{N} |x_i|^q \right)^{1/q}
\]

(13)

The \( L_{1/2} \) regularization is a non-convex, non-smooth and non-Lipschitz optimization problem. An iterative half thresholding algorithm for fast solution of \( L_{1/2} \) regularization was proposed and its convergence property was verified in [19], which corresponds to the well known iterative soft thresholding algorithm for \( L_1 \) regularization.

It is noted that GMR seeks to find a new mixture with as few number of components as possible to approximate the original mixture. If we can construct a Gaussian base set \( M \) with each component being the candidate of the reduced mixture, then we will be in the position to select as few components as possible from the base set and assign appropriate weights to approximate the original mixture. Obviously, it can be considered as a sparse modeling problem, which tries to find an appropriate weight \( w \) to minimize the error between the linear combination \( Mw \) and some representation \( y \) of original mixture \( \Omega_N \) with the sparse constraint \( \| w \|_0 = K \). That is

\[
\min_{x \in \mathbb{R}^n} \left\{ \| y - Mw \|^2 + \lambda \| w \|_{1/2}^{1/2} \right\}
\]

(14)

The following work will focus on how to construct such a Gaussian base \( M \) and model for the vector \( y \) and \( w \).

III GMR BASED ON \( L_{1/2} \) REGULARIZATION

A. Construction of Gaussian Base Set

In this subsection, we focus on constructing a Gaussian base set \( M \) by merging partial components from original Gaussian mixture. The key step is to develop a distance measure between two Gaussian mixtures and determine which components will be merged according to this measure.

It is obvious that we can use a greedy method to construct the base. Each element in the base set is produced by choosing \( l (l = 1, 2, ..., N) \) components by enumeration from \( \Omega_N \) and merging them. However, such a procedure will produce quite a large base set with the number of elements \( \sum_{l=1}^{N} C_N^l \), where \( C_N^l \) means the number of combinations of selecting \( l \) from \( N \). For example, if the component number \( N \) of original mixture \( \Omega_N \) equals to be 10, then the length of the base set will achieve 1023.

We noticed that such a greedy approach is infeasible and unnecessary, since those components with a larger dissimilarity should not be merged. What we need to do next is to find the \( k \) adjacent components \( U_{i,k}^* \) ( \( k = 1, 2, ..., N \) ) for any given component \( i (i = 1, 2, ..., N) \), and merge the components in \( U_{i,k}^* \) and add the new merged component to the base set \( M \).

Given any subset \( S_j \) of original mixture \( S_j \subseteq \Omega_N \), we denote the new component as \( M_{S_j} \) by merging all the components in \( S_j \). By the rule of preserving the moments of the overall mixture [12], the parameters of
new component \( M_{S_j} = \{w_j^m, \mu_j^m, P_j^m\} \) can be computed as
\[
w_j^m = \sum_{k \in S_j} w_k \tag{15}
\]
\[
\mu_j^m = \frac{1}{w_j^m} \sum_{k \in S_j} w_k \mu_k \tag{16}
\]
\[
P_j^m = \sum_{k \in S_j} \frac{w_k}{w_j^m} \left( P_k + (\mu_k - \mu_j^m)(\mu_k - \mu_j^m)^T \right) \tag{17}
\]

We define the cost of merging all the components in the subset \( S_j \) as \( c(S_j) \), by measuring the dissimilarity between two Gaussian mixtures \( \Omega_N \) and \( \{\Omega_N \setminus S_j, M_{S_j}\} \). In this paper, we use the ISD criterion, that is
\[
c(S_j) = \int_x \left( f(x|\Omega_N) - f\left(x|\{\Omega_N \setminus S_j, M_{S_j}\}\right) \right)^2 dx \tag{18}
\]

Now, when given any component \( i \) in \( \Omega_N (i = 1, \ldots, N) \), we define the \( k \) adjacent \( U_{i,k}^* \) of component \( i \) as a label set
\[
U_{i,k}^* = \arg \min_{|S_j|+k, \sum_{i \in S_j} c(S_j)} c(S_j) \tag{19}
\]

The corresponding cost of merging \( k \) adjacent components of component \( i \) is \( c(U_{i,k}^*) \). It is obvious that if \( k = 1 \), then \( k \) adjacent of component \( i \) is itself, ie \( U_{i,1}^* = \{w_i, \mu_i, P_i\} \). All the new components \( M_{U_{i,k}^*} \) \((i,k = 1,\ldots,N)\) will form the base set \( M \).

In addition, we introduce another label set \( B \) to record the component combinations, which corresponds to each element in the Gaussian base set \( M \). The details of constructing the Gaussian base set \( M \) and the label set \( B \) are given as follows.

**Step 1:** Initialize the Gaussian base set \( M = \Phi \), and the label set \( B = \Phi \).

**Step 2:** Given any component \( i \) \((i = 1,\ldots,N)\), compute the \( k \) adjacent components \( U_{i,k}^* \) \((k = 1,\ldots,N)\) and merge \( U_{i,k}^* \) into one component \( M_{U_{i,k}^*} = \{w_{i,k}^*, \mu_{i,k}^*, P_{i,k}^*\} \) by the rule of keeping the moment.

**Step 3:** If \( U_{i,k}^* \not\subseteq B \), then add \( U_{i,k}^* \) into the label set \( B = \{B \cup U_{i,k}^*\} \) and add \( M_{U_{i,k}^*} \) into the Gaussian base set \( M = \{M \cup M_{U_{i,k}^*}\} \).

To simplify the notation, we denote the final Gaussian base set as \( M = \{M_1, M_2, \ldots, M_L\}\), where each element \( M_j = \{\overline{w}_j, \overline{\mu}_j, \overline{P}_j\}\) \((j = 1,2,\ldots,L)\) corresponds to some \( M_{U_{i,k}^*} \) \((i,k = 1,\ldots,N)\) which is produced by merging all the components in \( U_{i,k}^* \). It can be easily seen that the number of elements of the Gaussian base set \( M \) constructed by the above three steps has an upper bound \( N^2 - N + 1 \). When the component number \( N \) of original mixture is large, the upper bound will be far less than the element number of \( 2^N - 1 \) if a base set is constructed by the greedy method we mentioned at the beginning of this subsection.

Once the Gaussian base set \( M \) is ready, what we need to do next is to choose \( K \)-components from the base set to approximate the original Gaussian mixture.

**B. \( L_{1/2} \) Regularization Formulation for GMR**

After having a Gaussian base set \( M \), the objective function (8) to optimize is turned into
where
\[ N_j(x) = \mathcal{N}(x; \mu_j, P_j) = \frac{1}{(2\pi)^{n/2} |P_j|^{1/2}} e^{-\frac{1}{2}(x-\mu_j)^T P_j^{-1} (x-\mu_j)} \]  
(21)

By noting that the integration of the product of two Gaussian PDFs,
\[ \int \mathcal{N}(x; \mu_1, P_1) \mathcal{N}(x; \mu_2, P_2) \, dx = \mathcal{N}(\mu; \mu_1, P_1 + P_2) \]  
(22)
The objective function \( J \) can be expressed as
\[ J = J_{NN} + J_{LL} - 2J_{NL} \]  
(23)
where
\[ J_{NN} \triangleq \int_{-\infty}^{\infty} \left[ f(x | \Omega_n) \right]^2 \, dx = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \mathcal{N} \left( \mu_i; \mu_j, P_i + P_j \right) \]  
(24)
and
\[ J_{LL} \triangleq \int_{-\infty}^{\infty} \left[ f(x | \Omega_L) \right]^2 \, dx = \sum_{i=1}^{N} \sum_{j=1}^{N} \eta_i \eta_j \mathcal{N} \left( \mu_i; \mu_j, P_i + P_j \right) \]  
(25)
and
\[ J_{NL} \triangleq \int_{-\infty}^{\infty} f(x | \Omega_N) \times f(x | \Omega_L) \, dx = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \eta_j \mathcal{N} \left( \mu_i; \mu_j, P_i + P_j \right) \]  
(26)
Considering that \( J_{NN} \) is not relevant to the parameter \( \eta = [\eta_1, \eta_2, \ldots, \eta_L]^T \), the minimizing of \( J \) is equivalent to minimize the following \( \tilde{J} \).
\[ \tilde{J} \triangleq J_{LL} - 2J_{NL} = \sum_{i=1}^{N} \sum_{j=1}^{N} Q_{ij}(\eta_i, \eta_j) - 2\sum_{j=1}^{L} y_j \eta_j \]  
(27)
where
\[ Q_{ij} \triangleq \mathcal{N} \left( \mu_i; \mu_j, P_i + P_j \right) \]  
(28)
and
\[ y_j \triangleq \sum_{i=1}^{N} w_i \mathcal{N} \left( \mu_i; \bar{\mu}_j, \bar{P}_i + \bar{P}_j \right) \]  
(29)

Further, if we denote \( Q = \{Q_{ij}\}_{i=1}^{N} \) and introduce \( \tilde{Q} = A^T A, y = A^T b \), then we have the following equivalent relationship
\[ \min_{\eta} \tilde{J} \Leftrightarrow \min_{\eta} \left( \| 4\eta - b \|_2 \right) \]  
(30)
So far, the following sparse modeling problem in the \( L_{1/2} \) framework is obtained.
\[ \min_{\eta} \left( \| 4\eta - b \|_2 + \lambda \| \eta \|_{1/2} \right) \]  
(31)
subject to
\[ \sum_{j=1}^{L} \eta_j = 1, \eta_j \geq 0, \text{supp}(\eta) = K \]  
(32)
where \( \text{supp}(\eta) \) is the support set of the vector \( \eta \) defined by
\[ \text{supp}(\eta) = \{ j \in 1, 2, \ldots, N, \eta_j \neq 0 \} \]  
(33)
The above optimization model (31-32) can be solved by the so-called iterative half thresholding algorithm given in [19]. For more details, we refer the reader to [19].

IV SIMULATION RESULTS

In this section, we provide a random scenario to validate the effectiveness and efficiency of the proposed \( L_{1/2} \) regularization solution for GMR. To better visualize the results of performance comparison, we only provide the scalar scenario. Obviously, the proposed method can also accommodate the scenario of high dimension. In this scenario, the component number \( N \) of the original mixture is set to be 10. The other parameters are designed as follows.

The mixing weight \( w_i \) samples from a Dirichlet distribution
\[
\mathbf{w}_i \sim \text{Dir}(n_1, \ldots, n_N) \tag{34}
\]
with the parameters
\[
n = [n_i]_{i=1}^{10} = [30 \ 20 \ 30 \ 40 \ 50 \ 50 \ 25 \ 20 \ 70 \ 40].
\]

The component mean \( \mu_i \) samples from a uniformly distribution
\[
\mathbf{\mu}_i \sim U(a, b) \tag{35}
\]
with the starting and ending points \( a = -5, b = 3 \).

The variance \( \sigma_i^2 \) for each Gaussian component samples from the inverse Gamma distribution
\[
\mathbf{\sigma}_i^2 \sim \text{IGamma}(\alpha, \beta) \tag{36}
\]
with the parameters \( \alpha = 50, \beta = 200 \).

We choose two algorithms for the performance comparison given in [2] and [13], called Salmond’s and GMRC, respectively. The initial value for GMRC is obtained by running Runnall’s algorithm in [8].

Our \( L_{1/2} \) regularization approach adopts a random chosen initial value.

Fig 1 shows the density function of the original Gaussian mixture and all its components. Fig 2-3 shows the performance of the proposed approach and competitive algorithms according to different number \( K \) of components of the reduced mixture. Fig 4 summarizes the GMR result of the proposed approach according to different component number \( K \). Table 1 summarizes ISD measure defined by
\[
\text{ISD} = \frac{1}{n} \left( \sum_{j=1}^{K} w_j \mathbf{N}_j(x) - \sum_{j=1}^{K} \tilde{w}_j \tilde{N}_j(x) \right)^2 \tag{37}
\]
Fig. 4. performance comparison with different number K of components of reduced mixture

Table.1. the ISD error for scenario 1

<table>
<thead>
<tr>
<th>Component number of reduced mixture</th>
<th>Salmond’s</th>
<th>GMC</th>
<th>Proposed approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>K=3</td>
<td>0.0465</td>
<td>0.0486</td>
<td>0.0168</td>
</tr>
<tr>
<td>K=2</td>
<td>0.0528</td>
<td>0.0889</td>
<td>0.0459</td>
</tr>
<tr>
<td>K=1</td>
<td>0.1677</td>
<td>0.1677</td>
<td>0.1677</td>
</tr>
</tbody>
</table>

V CONCLUSIONS

In this paper, a reduced Gaussian mixture representation based on sparse modeling is proposed. Simulation results demonstrated the better performance of the proposed approach compared with the competitive algorithms. Considering of the complexity and importance of constructing a Gaussian base set, future work includes designing a good distance criterion and a more efficient scheme to merge similar components. In addition, we can also plan to design a random Gaussian base set based on sampling from the original mixture to explore its sparse structure better.

ACKNOWLEDGEMENT

The work is sponsored by the national key fundamental research & development programs (973) of P.R. China (2007CB311006) and National Natural Science Foundation of China (61004087/F030119).

REFERENCES