Technical report from Automatic Control at Linköpings universitet

On Reduction of Mixtures of the Exponential Family Distributions

Tohid Ardeshiri, Emre Özkan , Umut Orguner

Division of Automatic Control

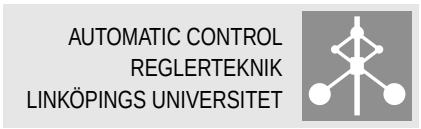
E-mail: tohid.ardeshiri@isy.liu.se, emre@isy.liu.se, umut@metu.edu.tr

18th August 2014

Report no.: LiTH-ISY-R-3076

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se



Abstract

Many estimation problems require a mixture reduction algorithm with which an increasing number of mixture components are reduced to a tractable level. In this technical report a discussion on different aspects of mixture reduction is given followed by a presentation of numerical simulation on reduction of mixture densities where the component density belongs to the exponential family of distributions.

Keywords: Mixture density, mixture reduction, exponential family, integral square error, Kullback-Leibler divergence, Exponential Distribution, Weibull Distribution, Laplace Distribution, Rayleigh Distribution, Log-normal Distribution, Gamma Distribution, Inverse Gamma Distribution, Univariate Gaussian Distribution, Multivariate Gaussian Distribution, Gaussian Gamma Distribution, Dirichlet distribution, Wishart Distribution, Inverse Wishart Distribution, Gaussian Inverse Wishart Distribution.

On Reduction of Mixtures of the Exponential Family Distributions

Tohid Ardeshiri, Emre Özkan and Umut Orguner

Abstract

Many estimation problems require a mixture reduction algorithm with which an increasing number of mixture components are reduced to a tractable level. In this technical report a discussion on different aspects of mixture reduction is given followed by a presentation of numerical simulation on reduction of mixture densities where the component density belongs to the exponential family of distributions.

Index Terms

Mixture density, mixture reduction, exponential family, integral square error, Kullback-Leibler divergence, Exponential Distribution, Weibull Distribution, Laplace Distribution, Rayleigh Distribution, Log-normal Distribution, Gamma Distribution , Inverse Gamma Distribution, Univariate Gaussian Distribution, Multivariate Gaussian Distribution, Gaussian Gamma Distribution, Dirichlet distribution, Wishart Distribution, Inverse Wishart Distribution, Gaussian Inverse Wishart Distribution

I. INTRODUCTION

A common problem encountered in statistical signal processing and tracking is mixture reduction (MR). Examples of such circumstances are multi-hypotheses tracking (MHT) [1], Gaussian sum filter [2], multiple model filtering [1], Gaussian mixture probability hypothesis density (GMPHD) filter [3]. In these algorithms the information about the state of a random variable is modeled as a mixture density. To be able to implement these algorithms for real time applications a mixture reduction step is necessary. The aim of the reduction algorithm is to reduce the computational complexity into a predefined budget while keeping the inevitable error introduced by the approximation as small as possible.

The problem of reducing a mixture density to another mixture density with less components is addressed in several papers such as [4]–[13]. In [4] and [5], a similarity measure is used for merging the components of a mixture. Covariance union and generalized covariance union are described in [6] and [7], respectively. A Gaussian MR algorithm using clustering techniques is suggested by [10]. A Gaussian MR algorithm using homotopy to avoid local minima is suggested by [9]. [8] summarizes the Gaussian MR algorithms and compares them. Merging statistics for greedy MR for multiple target tracking is discussed in [13]. Most of these papers are focused on reduction of Gaussian mixtures due to widespread use of Gaussian density and Kalman filter in estimation problems.

Two of the recent algorithms which are most related to this work are [14] and [15]. The Kullback-Leibler approach to Gaussian MR is introduced by Runnalls in [15]. In this work Runnalls uses an upper bound on the Kullback-Leibler divergence (KLD) as a measure of similarity between the original mixture density and its reduced form in each step of reduction. The motivation for the choice of an upper bound in Runnalls's algorithm is that the KLD between two Gaussian mixtures can not be calculated analytically. The method of [15] is compared to other reduction methods in [8] and it is suggested to be the most practical Gaussian MR algorithm for target tracking by then. The Integral Square Error (ISE) approach for MR was used by Williams and Maybeck in [14]. The great advantage of the ISE approach is that it has a closed form solution for the similarity measure it proposes between two Gaussian mixtures.

The reduction problem can be formulated as a nonlinear optimization problem where cost functions such as Kullback-Leibler divergence or integral squared error [16] are selected and the optimization is solved by numerical solvers. In these approaches the number of components in the reduced Gaussian mixture can be known in advance or not. These optimization approaches can be quite expensive especially for high dimensional data and not suitable for real time implementation.

In this technical report the mixture reduction (MR) problem is cast as a decision problem and the maximum a posteriori (MAP) decision rule is derived for the mixture reduction problem. Also, four variants of integral square error (ISE) approach for the decision problem are discussed and a computationally cost saving scheme is given in Section II-E. In Section III a case study on exponential family of distributions is performed where a merging algorithm for exponential family of distributions is derived. Also, an example on reduction of Gaussian Inverse Wishart mixture density is presented. In Section III-B the Monte Carlo (MC) simulation scenario and its results are given for some common members of the exponential family of distributions. In Appendix B expressions needed for solving each the mixture reduction problem for common members of the exponential family of distributions are provided.

A. Essentials of greedy MR algorithms

A *mixture density* is a probability density which is a convex combination of other (more basic) probability densities [17]. For example, consider a normalized mixture consisting of N components given by

$$p(x) = \sum_{I=1}^{N} w^I q(x; \eta^I), \tag{1}$$

T. Ardeshiri and E. Özkan are with the Department of Electrical Engineering, Linköping University, 58183 Linköping, Sweden, e-mail: tohid,emre@isy.liu.se.
U. Orguner is with Department of Electrical and Electronics Engineering, Middle East Technical University, 06531 Ankara Turkey, email: umut@metu.edu.tr
.

where I is a categorical variable with $p(I) = w^I$. The parameters of the density $q(x;\eta^I)$ are comprised in η^I. Now, consider the reduction of the mixture (1) to another mixture consisting of $\mathbf{M}$ components where $1 < \mathbf{M} < N$. In the following, two operations which are used in the reduction of number of components from N components to $\mathbf{M}$ components, namely pruning and merging, are described.

Pruning which is the simplest approach to reducing the number of components in a mixture density is to remove one (or more) components of a mixture and rescaling the remaining components such that it integrates to unity. For example pruning component J from (1) results in the mixture density $(1-w^J)^{-1}\sum_{I=1,I\neq J}^{N} w^I q(x;\eta^I)$. The error introduced at point x belonging to support of the density by pruning a component in (1) with label J obeys

$$\begin{aligned} e^{0J}(x,p(x)) &= p(x) - (1-w^J)^{-1}\sum_{I=1,I\neq J}^{N} w^I q(x;\eta^I) \\ &= w^J(1-w^J)^{-1}q(x;\eta^J) - w^J(1-w^J)^{-1}p(x). \end{aligned} \tag{2}$$

The approximation of a normalized partition of a mixture by a single component via minimizing a similarity measure, such as the Kullback-Leibler divergence (KLD), between the partition and the single component is referred to as *merging*. For example, consider the problem of merging a subset of components with labels $\mathcal{L} \subset \{I\}_{I=1}^{N}$ into one component. A merging algorithm can be proposed for approximating a mixture density given as

$$p_{\mathcal{L}}(x) = \sum_{I\in\mathcal{L}} \widehat{w}^I q(x;\eta^I), \tag{3}$$

where $\widehat{w}^I = (\sum_{j\in\mathcal{L}} w^j)^{-1}(w^I)$, by a single component $q(x;\eta^L)$ via minimizing the KLD between the densities $p_{\mathcal{L}}(x)$ and $q(x;\eta^L)$, where the KLD is defined by $D_{KL}(p_{\mathcal{L}}||q_L) = \int p_{\mathcal{L}} \log \frac{p_{\mathcal{L}}}{q_L} dx$ and q_L is short notation for $q(x;\eta^L)$. In other words, the optimal parameter η^{L*} is given by

$$\eta^{L*} = \underset{\eta^L}{\operatorname{argmin}}\, D_{KL}(p_{\mathcal{L}}||q_L). \tag{4}$$

For more explicit expressions for merging mixture densities belonging to the exponential family via KLD minimization see [18]. The error introduced in each step of merging two components I and J of the mixture density (1) is

$$e^{IJ}(x) = w^I q(x;\eta^I) + w^J q(x;\eta^J) - w^{IJ} q(x;\eta^{IJ}), \tag{5}$$

where $w^{IJ} = w^I + w^J$ and η^{IJ} is the parameter of the component resulting from approximation of the two components I and J.

A *greedy* approach to MR is to reduce the mixture to a mixture composed of less components via one to one comparison of the components and merging selected pairs or pruning selected components. Such MR methods have two elements; a metric for comparing different hypotheses and a merging algorithm. The metric is used to compare a merging or pruning hypothesis to another merging or pruning hypothesis. If a hypothesis of merging is preferred, the selected pair of components are merged into a single new component. In the context of greedy MR we need to define a paradigm to decide which components to prune and which components to merge. In the following section three of such paradigms namely, approximate Kullback-Leibler, Integral Square Error and Symmetrized Kullback-Leibler are introduced.

II. Mixture reduction as a decision problem

The reduction problem can be formulated as a decision problem where most likely hypothesis regarding the distribution of the data can be selected. Before discussing the decision metrics in sections II-B, II-D and II-F, let us have a look at the possible decisions at each stage of the reduction.

At the k^{th} stage of reducing mixture density of equation (1), $n_k = N-k+1$ components are left and there are $\frac{1}{2}n_k \times (n_k-1)$ possible merging decisions and n_k possible pruning decisions to choose from. Let the reduced density at the k^{th} stage be denoted by $p_k(x)$. We have a multiple hypotheses decision problem at hand where the hypotheses are formulated according to

$$\text{Pruning Hypotheses} \begin{cases} \mathcal{H}_{01} : p_k(x|\mathcal{H}_{01}) = p_k(x) - e^{01}(x,p_k(x)), \\ \mathcal{H}_{02} : p_k(x|\mathcal{H}_{02}) = p_k(x) - e^{02}(x,p_k(x)), \\ \qquad\qquad \vdots \\ \mathcal{H}_{0n_k} : p_k(x|\mathcal{H}_{0n_k}) = p_k(x) - e^{0n_k}(x,p_k(x)), \end{cases}$$

$$\text{Merging Hypotheses} \begin{cases} \mathcal{H}_{12} : p_k(x|\mathcal{H}_{12}) = p_k(x) - e^{12}(x), \\ \mathcal{H}_{13} : p_k(x|\mathcal{H}_{13}) = p_k(x) - e^{13}(x), \\ \qquad\qquad \vdots \\ \mathcal{H}_{(n_k-1)n_k} : p_k(x|\mathcal{H}_{(n_k-1)n_k}) = p_k(x) - e^{(n_k-1)n_k}(x), \end{cases}$$

which is a decision problem with $n_k(n_k+1)/2$ hypotheses. The first n_k hypotheses account for pruning and the rest account for merging decisions. The subscript on hypotheses $\mathcal{H}_{\cdot}$ refers to the two components to be merged for merging hypotheses

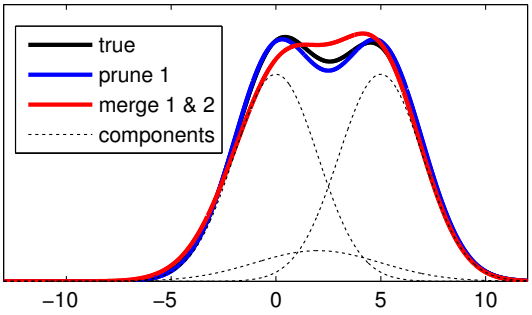


Figure 1. A Gaussian mixture is compared with two approximation via pruning component 1 (in blue) and merging components 1 and 2 (in red)

.

while in the case of pruning hypotheses the subscript refers to the label of the component to be pruned which is preceded by zero.

The number of merging hypotheses which needs to be evaluated for the reduction of a mixture with N components down to $M = N - k$ components is $\frac{1}{2}N^2k + \frac{1}{6}k^3 - \frac{1}{2}Nk^2 - \frac{1}{6}k$. Also, the number of pruning hypotheses which needs to be evaluated for the reduction of a mixture with N components down to $N - k$ components is $Nk + \frac{1}{2}N - \frac{1}{2}k^2$.

The basic problem in statistical decision theory is to make optimal choices from a set of alternatives, i.e., hypotheses, based on noisy observations [19]. The intention is here to cast the problem of greedy mixture reduction for a given merging and pruning algorithm in the framework of statistical decision theory. The alternative decisions to be made in the reduction problem is which components to merge or prune. However, there are no observation to base the decision on. The remedy is to use the original mixture $p(x)$ or more precisely infinitely many hypothetical independent and identically distributed (iid) samples from the original mixture as the observations i.e., $\{x^r\}_{r=1}^{\infty} \overset{\text{iid}}{\sim} p(x)$. It should be emphasized that the samples are hypothetical and later on using the law of large numbers the need for them will be eliminated.

A. Hypothesis testing for greedy MR

Using the multiple hypothesis testing framework and the data X, the hypothesis $\mathcal{H}_K$ with $P(\mathcal{H}_K|X) > P(\mathcal{H}_I|X)$ for $I \neq K$ should be decided, see Appendix A or [19]. This decision rule is also referred to as maximum a posteriori (MAP) decision rule. For equal prior probabilities $P(\mathcal{H}_K) = P(\mathcal{H}_I)$ the decision rule will be to decide $\mathcal{H}_K$ if $p(X|\mathcal{H}_K) > p(X|\mathcal{H}_I)$ for $I \neq K$. Now suppose the data $\{x^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p(x)$ is used for the decision problem of deciding which hypothesis to choose among merging and pruning decisions. Therefore the likelihood can be written as

$$p(\{x^r\}_{r=1}^{S}|\mathcal{H}_K) = \prod_{r=1}^{S} p(x^r|\mathcal{H}_K) \tag{6}$$

with the assumption that the samples $\{x^r\}_{r=1}^{S}$ are independent. The normalized log-likelihood by dividing by the number of observations S, is given by $\widehat{l}(\{x^r\}_{r=1}^{S}|\mathcal{H}_K) \triangleq \frac{1}{S}\sum_{r=1}^{S} \log p(x^r|\mathcal{H}_K)$. From the strong law of large numbers we have that $\widehat{l}(\{x^r\}_{r=1}^{S}|\mathcal{H}_K) \overset{a.s.}{\to} \mathbb{E}_{p(x)}[\log p(x|\mathcal{H}_K)]$. Therefore the optimal decision is to decide $\mathcal{H}_K$ if

$$\mathop{\mathbb{E}}_{p(x)}[\log p(x|\mathcal{H}_K)] > \mathop{\mathbb{E}}_{p(x)}[\log p(x|\mathcal{H}_I)]$$

for $K \neq I$. It can be shown by simple calculations that this decision criterion is equivalent to decide $\mathcal{H}_K$ if

$$D_{KL}(p(x)||p(x|\mathcal{H}_K)) < D_{KL}(p(x)||p(x|\mathcal{H}_I)) \tag{7}$$

for $K \neq I$. The MAP decision rule (7) will be used in Example 1 to decide between two hypotheses.

Example 1: Consider a Gaussian mixture consisting of three components $p(x) = \sum_{I=1}^{3} w^I \mathcal{N}(x; m^I, P^I)$ as illustrated in Figure 1. The parameters of each component, weight w^I, mean value m^I and variance P^I, are given as $w^1 = 0.1$, $m^1 = 2$, $P^1 = 9$, $w^2 = 0.45$, $m^2 = 0$, $P^2 = 4$, $w^3 = 0.45$, $m^3 = 5$, $P^3 = 4$. In order to compare the merging hypothesis $\mathcal{H}_{12}$ and the pruning hypothesis $\mathcal{H}_{01}$ using the MAP rule given in (7) we calculate the KLD between the original mixture and the candidate approximations which gives $D_{KL}(p(x)||p(x|\mathcal{H}_{01})) = 0.00053$ and $D_{KL}(p(x)||p(x|\mathcal{H}_{12})) = 0.0026$. Therefore the MAP decision rule favors pruning component 1 for reducing the mixture to 2 components. ■

B. Approximate Kullback-Leibler (AKL) approach to mixture reduction

The MAP decision rule of (7) requires the computation of KLD between two mixture densities. Unfortunately there is no analytical expression for the Kullback-Leibler divergence of two mixtures of the same member of commonly used basic densities except for Categorical distribution. In [20], several methods for the approximation of the KLD between Gaussian

mixtures, which include Monte Carlo sampling and Unscented transform, are given which can be generalized for other mixture densities.

An upper bound for the KLD of two mixtures suggested in [15] is the most promising of all approximate KLD methods for reduction of Gaussian mixtures for target tracking. It is evaluated in [8] and suggested to be the most practical Gaussian mixture reduction to use in a tracker thus far. The approximate KLD based decision criteria introduced in [15] can be given in more general form using [15, Theorem 3 and Theorem 4] which will be given here for completeness.

Theorem 1. *If $f(x)$, $h_1(x)$ and $h_2(x)$ are any probability distribution functions and $0 \leq w \leq 1$ then*

$$D_{KL}(wh_1(x) + (1-w)h_2(x)||f(x)) \leq wD_{KL}(h_1(x)||f(x)) + (1-w)D_{KL}(h_2(x)||f(x))$$
$$D_{KL}(f(x)||wh_1(x) + (1-w)h_2(x)) \leq wD_{KL}(f(x)||h_1(x)) + (1-w)D_{KL}(f(x)||h_2(x)).$$

■

Theorem 2. *If $f(x)$, $h_1(x)$ and $h_2(x)$ are any probability distribution functions and $0 \leq w \leq 1$ then*

$$D_{KL}(wh_1(x) + (1-w)f(x)||wh_2(x) + (1-w)f(x)) \leq wD_{KL}(h_1(x)||h_2(x))$$

■

Using theorems 1 and 2, $D_{KL}(p(x)||p(x|\mathcal{H}_{IJ}))$ is approximated in [15] for the case of merging two components I and J as in

$$\begin{aligned} &D_{KL}(p(x)||p(x|\mathcal{H}_{IJ})) \\ &\leq (w^I + w^J)D_{KL}\left(\frac{w^I q(x;\eta^I) + w^J q(x;\eta^J)}{w^I + w^J}\middle|\middle| q(x;\eta^{IJ})\right) \\ &\leq w^I D_{KL}(q(x;\eta^I)||q(x;\eta^{IJ})) + w^J D_{KL}(q(x;\eta^J)||q(x;\eta^{IJ})). \end{aligned} \tag{8}$$

The upper bound given above on $D_{KL}(p(x)||p(x|\mathcal{H}_{IJ}))$ is defined by

$$\mathcal{B}(I,J) \triangleq w^I D_{KL}(q(x;\eta^I)||q(x;\eta^{IJ})) + w^J D_{KL}(q(x;\eta^J)||q(x;\eta^{IJ})) \tag{9}$$

The weighted sum in (9) will be referred to as $\mathcal{B}(I,J)$ for the general case, as in the original formulation for Gaussian mixtures in [15]. The advantages of using $\mathcal{B}(I,J)$ instead of $D_{KL}(p(x)||p(x|\mathcal{H}_{IJ}))$ is that $D_{KL}(q(x;\eta^I)||q(x;\eta^J))$ has a closed form expression for many densities of interest.

The AKL decision criteria is a very strong tool in comparison of merging hypotheses. In [15], no counterpart for $\mathcal{B}(I,J)$ is given for comparison of the pruning hypotheses. In the following, a pruning metric for the pruning hypotheses using Theorems 1 and 2 will be derived.

C. Pruning a mixture based on approximate KLD

The KLD of the two mixtures $p(x)$ and $p(x|\mathcal{H}_{0I})$ which arise when computing the decision rule (7) for hypothesis of pruning component I can be approximated using theorems 1 and 2 as in

$$\begin{aligned} &D_{KL}(p(x)||p(x|\mathcal{H}_{0I})) = D_{KL}(w^I q_{\eta^I} + (1-w^I)r||w^I r + (1-w^I)r) \leq w^I D_{KL}(q_{\eta^I}||r) \\ &\leq \frac{w^I}{1-w^I}\sum_{j=1,j\neq I}^{N} w^j D_{KL}(q_{\eta^I}||q_{\eta^j}), \end{aligned} \tag{10}$$

where $r = \frac{1}{1-w^I}\sum_{j=1,j\neq I}^{N} w^j q_{\eta^j}$. The upper bound on $D_{KL}(p(x)||p(x|\mathcal{H}_{0I}))$ becomes

$$\mathcal{B}(0,I) = \frac{w^I}{1-w^I}\sum_{j=1,j\neq I}^{N} w^j D_{KL}(q_{\eta^I}||q_{\eta^j}). \tag{11}$$

According to (11), a small value for $\mathcal{B}(0,I)$ can be obtained when the KLD between the component I and all the remaining components $\{j\}_{j=1,j\neq I}^{N}$ is small which is an unlikely scenario. Furthermore, if component I is similar to all the other components $\{j\}_{j=1,j\neq I}^{N}$ merging of component I with one of these remaining components, $\{j\}_{j=1,j\neq I}^{N}$, should be a good alternative to pruning component I. Based on the authors' experience this KLD based pruning metric has shown no practical use in many numerical experiments conducted by the authors.

Table I
PARAMETERS OF THE GAUSSIAN MIXTURE OF FIGURE 2 WITH 3 COMPONENTS ARE GIVEN.

I	w^I	m^I	P^I
1	0.1	-2	16
2	0.45	0	1
3	0.45	5	1

Table II
ISE AND KLD FOR RESULTING MIXTURE FROM TWO HYPOTHESES OF MERGING ($K = 12$) AND PRUNING ($K = 01$) ARE GIVEN FOR COMPARISON.

K	$D_{KL}(p(x)\|\|p(x\|\mathcal{H}_K))$	$\mathrm{ISE}(\mathcal{H}_K) \times 10^{-3}$
01	0.6134	1.0293
12	0.1328	12.961

D. Integral Square Error (ISE) approach

An alternative solution for the multiple hypothesis decision problem which arises in the MR is the ISE approach [14], [18]. ISE is a measure of difference between two densities which is defined as $\int |f(x) - g(x)|^2 \, \mathrm{d}x$ for two densities $f(x)$ and $g(x)$. ISE is used by Williams and Maybeck in [14] where a cost is associated with each reduction hypothesis. The cost of the hypothesis $\mathcal{H}_K$ obeys

$$\mathrm{ISE}(\mathcal{H}_K) = \int |p(x) - p_k(x|\mathcal{H}_K)|^2 \, \mathrm{d}x. \tag{12}$$

In this approach, the hypothesis which gives the smallest ISE will be chosen at each step of the reduction i.e., the decision rule based on ISE becomes "decide $\mathcal{H}_K$ if $\mathrm{ISE}(\mathcal{H}_K) < \mathrm{ISE}(\mathcal{H}_L)$ for all $L \neq K$", where K and L are permissible indices of the hypotheses. For the Gaussian mixture of Example 1, $\mathrm{ISE}(\mathcal{H}_{01}) = 0.000062$ and $\mathrm{ISE}(\mathcal{H}_{12}) = 0.00029$ i.e., ISE decision rule favors pruning component 1 for the reduction of the mixture to 2 components. Another comparison of the ISE decision rule and the MAP decision rule will be given in the following example.

Example 2: Consider a Gaussian mixture consisting of three components; $p(x) = \sum_{I=1}^{3} w^I \mathcal{N}(x; m^I, P^I)$. The parameters of each component, mean value m^I and variance P^I are given in Table I. Now, in order to compare the merging decision $\mathcal{H}_{12}$ and the pruning decision $\mathcal{H}_{01}$ using the ISE and the MAP criteria, these values are calculated numerically and are given in Table II. From the second column of Table II it can be inferred that the KLD between the original mixture and the mixture consisting of the third ($I = 3$) component and the merged component of 1 and 2 is lower than the KLD between the original mixture and the re-normalized mixture consisting of only the second and the third component. That is, the MAP decision rule favors merging decision. On the other hand, ISE approach favors pruning decision (see second column in Table II). The original mixture and the two approximations are shown in Figure 2 for illustration. ■

Examples 1 and 2 show the impact the choice of paradigm has on the outcome of the decision problem. The KLD between two densities $p(x)$ and $q(x)$ is given by $D_{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x$ while the ISE is defined by $\mathrm{ISE}(p, q) = \int |p(x) - q(x)|^2 \, \mathrm{d}x$. The MAP decision rule attempts to minimize the expected value of the logarithm of the "ratio" between the two probability distribution functions (pdf) over the support of the pdfs. In contrast ISE is mainly concerned about the "square of the difference" between the two densities. The former tends to infinity in the limit when the approximate density vanishes where the original density is nonzero while the latter is always finite. This explains why the MAP decision rule tries to avoid pruning while ISE approach has a tendency to prune the components with small weight.

An advantage of the ISE metric is that, it can be computed analytically for many distributions [18]. In the ISE approach two parameters can be varied to create slightly different reduction algorithms as detailed below:

1) In the first variation, the ISE is calculated for each hypothesis according to $\mathrm{ISE}(\mathcal{H}_K) = \int |p(x) - p_k(x|\mathcal{H}_K)|^2 \, \mathrm{d}x$ and the density after pruning is re-normalized. This variation is consistent with the presentation of the ISE algorithm so far in this technical report.

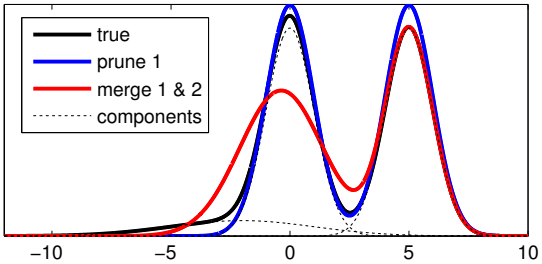


Figure 2. The original mixture with parameters given in Table I is compared with two approximation favored by ISE (in blue) and MAP (red).

2) In the second variation, as it is pointed out in [14], when the ISE is being calculated for a pruning hypothesis the rescaling can be skipped since re-normalizing the weights will increase the error value in parts of the support that are not affected by the pruning hypothesis. This choice also brings substantial computational saving which will be discussed is section II-E.
3) In the third variation, instead of comparing $p(x|\mathcal{H}_K)$ with the original mixture $p(x)$, it is compared with the resulting mixture of the previous reduction step $p_k(x)$, as given here

$$\mathrm{ISE}(\mathcal{H}_K) = \int |p_k(x) - p_k(x|\mathcal{H}_K)|^2 dx = \int |e^K|^2 \,\mathrm{d}x.$$

In this way, the ISE metric for merging decision can be simplified to

$$\begin{aligned}\mathrm{ISE}(\mathcal{H}_{IJ}) = {} & (w^I)^2 Q(I,I) + (w^J)^2 Q(J,J) \\ & + (w^{IJ})^2 Q(IJ,IJ) + 2w^I w^J Q(I,J) \\ & - 2w^I w^{IJ} Q(I,IJ) - 2w^J w^{IJ} Q(J,IJ).\end{aligned}$$

where,

$$Q(I,J) = \int q(x;\eta^I) q(x;\eta^J) \,\mathrm{d}x. \tag{13}$$

$Q(I,J)$ can be calculated analytically for many basic densities of interest belonging to the exponential family such as Gaussian, gamma and Wishart distributions. For explicit expressions for the exponential family of distribution see [18]. Similarly the ISE metric for pruning decision can be simplified as

$$\mathrm{ISE}(\mathcal{H}_{0I}) = \left(\frac{w^I}{1-w^I}\right)^2 \left[Q(I,I) - 2\sum_{i=1}^{N} w^i Q(I,i) + \sum_{i=1}^{N}\sum_{j=1}^{N} w^i w^j Q(i,j) \right].$$

4) The fourth variant is similar to the third variant in terms of the choice of the reference density, but the mixture is not renormalized after each pruning which results in the expression

$$\mathrm{ISE}(\mathcal{H}_{0I}) = (w^I)^2 Q(I,I)$$

for pruning hypotheses.

The second variation is used in the numerical simulation that is presented in section III-B.

E. Implementation aspects of the ISE approach

Calculation of the ISE for each hypothesis at every step of the reduction is costly. A scheme is suggested here to cache the calculated quantities to reduce the computational cost of the reduction. The cost reduction scheme is given for the second type of implementation of the ISE approach according to the numbered list in section II-D where the mixture density after pruning hypothesis is not re-normalized.

In the first step of the reduction of the mixture density (1) merging of all possible pairs of components results in $\frac{1}{2}N(N-1)$ hypotheses. For the evaluation of these hypotheses the resultant component of each merging should be calculated. To calculate the ISE of each hypothesis $Q(\cdot,\cdot)$ should be calculated for all pairs of components in the mixture as well as the pair of components where one component is among the merged components and the other one is among the existing components. All these quantities should be stored and can be reused in the future reduction steps.

At the k^{th} step of the reduction of the mixture density given in (1), the reduced density is denoted by $p_k(x)$. In order to keep the notation less cluttered, let the term q^J denote $w^J q(x;\eta^J)$; p denote $p(x)$ and p_k denote $p_k(x)$. Let us assume that the cost of the reduction hypotheses at the k^{th} stage denoted by $ISE_k(\mathcal{H}_R)$ are stored in a vector Y_k and let $M = \operatorname{argmin} ISE_k(\mathcal{H}_R)$ for all permissible values of R.

When M corresponds to a pruning hypothesis, for example $M = 0J$, the vector Y_{k+1} can be updated with less computations for next pruning hypotheses using

$$\begin{aligned}\mathrm{ISE}_{k+1}(\mathcal{H}_{0S}|M=0J) &= \int (p - p_k + q^J + q^S)^2 \,\mathrm{d}x \\ &= \int (p - p_k + q^S)^2 \,\mathrm{d}x + \int (q^J)^2 \,\mathrm{d}x + 2\int q^J (p - p_k + q^S) \,\mathrm{d}x \\ &= \int (p - p_k + q^S)^2 \,\mathrm{d}x + \int (q^J)^2 \,\mathrm{d}x + 2\int q^J (p - p_k) \,\mathrm{d}x + 2\int q^J q^S \,\mathrm{d}x \\ &= \mathrm{ISE}_k(\mathcal{H}_{0S}) + \underbrace{\int (q^J)^2 \,\mathrm{d}x + 2\int q^J (p - p_k) \,\mathrm{d}x}_{A(J)} + 2\int q^J q^S \,\mathrm{d}x,\end{aligned} \tag{14}$$

where, the quantity $\mathrm{ISE}_k(\mathcal{H}_{0S})$ is already known from the previous step and $A(J)$ is a part of the ISE added to elements of Y_k due to the pruning of the J^{th} component.

Similarly, when M corresponds to a pruning hypothesis, for example $M = 0J$, the vector Y_{k+1} can be updated with less computations for the next merging hypotheses using

$$\begin{aligned}
&\mathrm{ISE}_{k+1}(\mathcal{H}_{ST}|M = 0J) = \int (p - p_k + q^J + q^S + q^T - q^{ST})^2 \,\mathrm{d}x \\
&= \int (p - p_k + q^S + q^T - q^{ST})^2 \,\mathrm{d}x + \int (q^J)^2 \,\mathrm{d}x + 2\int q^J(p - p_k + q^S + q^T - q^{ST}) \,\mathrm{d}x \\
&= \int (p - p_k + q^S + q^T - q^{ST})^2 \,\mathrm{d}x + \int (q^J)^2 \,\mathrm{d}x + 2\int q^J(p - p_k) \,\mathrm{d}x + 2\int q^J(q^S + q^T - q^{ST}) \,\mathrm{d}x \\
&= \mathrm{ISE}_k(\mathcal{H}_{ST}) + A(J) + 2\int q^J(q^S + q^T - q^{ST}) \,\mathrm{d}x.
\end{aligned} \tag{15}$$

After each pruning step all elements of vector Y_{k+1} corresponding to the pruned component will be eliminated from Y_{k+1}.

Using a similar approach, when M corresponds to a merging hypothesis, say $M = IJ$, the vector Y_{k+1} can be updated with less computations for the next pruning hypotheses using

$$\begin{aligned}
&\mathrm{ISE}_{k+1}(\mathcal{H}_{0S}|M = IJ) = \int (p - p_k + q^J + q^I - q^{IJ} + q^S)^2 \,\mathrm{d}x \\
&= \int (p - p_k + q^S)^2 \,\mathrm{d}x + \int (q^J + q^I - q^{IJ})^2 \,\mathrm{d}x + 2\int (q^J + q^I - q^{IJ})(p - p_k + q^S) \,\mathrm{d}x \\
&= \int (p - p_k + q^S)^2 \,\mathrm{d}x + \int (q^J + q^I - q^{IJ})^2 \,\mathrm{d}x + 2\int (q^J + q^I - q^{IJ})(p - p_k) \,\mathrm{d}x + 2\int (q^J + q^I - q^{IJ})q^S \,\mathrm{d}x \\
&= \mathrm{ISE}_k(\mathcal{H}_{0S}) + \underbrace{\int (q^J + q^I - q^{IJ})^2 \,\mathrm{d}x + 2\int (q^J + q^I - q^{IJ})(p - p_k) \,\mathrm{d}x}_{C(I,J)} + 2\int (q^J + q^I - q^{IJ})q^S \,\mathrm{d}x,
\end{aligned} \tag{16}$$

and for the next merging hypotheses using

$$\begin{aligned}
&\mathrm{ISE}_{k+1}(\mathcal{H}_{ST}|M = IJ) = \int (p - p_k + q^J + q^I - q^{IJ} + q^S + q^T - q^{ST})^2 \,\mathrm{d}x \\
&= \int (p - p_k + q^S + q^T - q^{ST})^2 \,\mathrm{d}x + \int (q^J + q^I - q^{IJ})^2 \,\mathrm{d}x + 2\int (q^J + q^I - q^{IJ})(p - p_k + q^S + q^T - q^{ST}) \,\mathrm{d}x \\
&= \int (p - p_k + q^S + q^T - q^{ST})^2 \,\mathrm{d}x + \int (q^J + q^I - q^{IJ})^2 \,\mathrm{d}x + 2\int (q^J + q^I - q^{IJ})(p - p_k) \,\mathrm{d}x \\
&+ 2\int (q^J + q^I - q^{IJ})(q^S + q^T - q^{ST}) \,\mathrm{d}x \\
&= \mathrm{ISE}_k(\mathcal{H}_{ST}) + C(I,J) + 2\int (q^J + q^I - q^{IJ})(q^S + q^T - q^{ST}) \,\mathrm{d}x.
\end{aligned} \tag{17}$$

When two components I and J are merged, the merged component labeled IJ will obtain the label of component I in the computation environment and all elements of Y_{k+1} corresponding to element J will be eliminated. The vector Y_{k+1} should be updated for the new component as in

$$\begin{aligned}
&\mathrm{ISE}_{k+1}(\mathcal{H}_{(IJ)S}|M = IJ) = \int (p - p_k + q^J + q^I - q^{IJ} + q^S + q^{IJ} - q^{(IJ)S})^2 \,\mathrm{d}x \\
&= \int (p - p_k)^2 \,\mathrm{d}x + \int (q^J + q^I + q^S - q^{(IJ)S})^2 \,\mathrm{d}x + 2\int (p - p_k)(q^J + q^I + q^S - q^{(IJ)S}) \,\mathrm{d}x,
\end{aligned} \tag{18}$$

where, the first term is known from the last reduction step.

F. Symmetrized Kullback-Leibler Divergence (SKL) Approach

As another similarity measure the Symmetrized Kullback-Leibler Divergence is used for the comparison of the merging hypotheses in [21], [22], [23] and [24]. The symmetrized KLD (SKL) for two component densities is defined as

$$D_{SKL}(I, J) = D_{KL}(q_{\eta^I}||q_{\eta^J}) + D_{KL}(q_{\eta^J}||q_{\eta^I}). \tag{19}$$

This approach is used in the numerical simulation intended for comparison of different MR algorithms in section III-B. For further details on implementation see [18].

III. Case study on reduction of Exponential family of distributions

The exponential family in its natural form can be represented by its natural parameters η, sufficient statistic $T(x)$, Log-partition function $A(\eta)$ and base measure $h(x)$ as in

$$q(x;\eta) = h(x)\exp(\eta \cdot T(x) - A(\eta)), \tag{20}$$

where the natural parameter η belongs to the natural parameter space $\Omega = \{\eta \in \mathbb{R}^m | A(\eta) < +\infty\}$. Here $a \cdot b$ denotes the inner product of a and b. In the following we will write q_η and q_L as shorthand notations for $q(x;\eta)$ and $q(x;\eta^L)$ respectively to keep the notation less cluttered. Some properties of the log partition function $A(\eta)$ will be given in the following to be used later on.

Definition 1. *The set corresponding to all mean values for the sufficient statistics*

$$\mathcal{M} = \{\mu \in \mathbb{R}^m | \exists p, \mathop{\mathbb{E}}_{p}[T(x)] = \mu\} \tag{21}$$

is called the mean parameter space [25].

Definition 2. *In a regular family of exponential family the domain Ω is an open set [25].*

Definition 3. *In minimal representation of an exponential family a unique parameter vector is associated with each distribution [25].*

Lemma 1. *For $q(x;\eta)$ belonging to the exponential family, the following holds*

$$q(x;\eta)T(x) = \nabla_\eta q(x;\eta) + q(x;\eta)\nabla_\eta A(\eta). \tag{22}$$

Proof.

$$\begin{aligned}\nabla_\eta q(x;\eta) &= h(x)\exp(\eta \cdot T(x) - A(\eta))(T(x) - \nabla_\eta A(\eta)) \\ &= q(x;\eta)(T(x) - \nabla_\eta A(\eta)).\end{aligned}$$

and hence, (22) is achieved. □

Proposition 1. *The gradient of the log partition function $\nabla A : \Omega \to \mathcal{M}$ associated with any regular exponential family has derivatives of all orders on its domain and the first two derivatives yield the cumulants of the sufficient statistics $T(x)$. Moreover, A is a convex function of η on its domain Ω and strictly convex if the representation is minimal.*

Proof. Here the expressions for the first two derivatives are derived. For a complete proof see [25, Proposition 3.1].

$$\begin{aligned}\nabla_\eta A(\eta) &= \nabla_\eta A(\eta) + \nabla_\eta \int q_\eta \,\mathrm{d}x \\ &= \int q_\eta \nabla_\eta A(\eta) + \nabla_\eta q_\eta \,\mathrm{d}x = \int q_\eta T(x) \,\mathrm{d}x, \\ &= \mathop{\mathbb{E}}_{q_\eta}[T(x)].\end{aligned} \tag{23}$$

For the second derivative we have

$$\begin{aligned}\nabla_\eta^2 A(\eta) &= \nabla_\eta \mathop{\mathbb{E}}_{q_\eta}[T(x)] = \nabla_\eta \int q_\eta T(x) \,\mathrm{d}x \\ &= \int \big(\nabla_\eta q_\eta\big) T(x)^{\mathrm{T}} \,\mathrm{d}x \\ &= \int q_\eta (T(x) - \nabla_\eta A(\eta)) T(x)^{\mathrm{T}} \,\mathrm{d}x \\ &= \mathop{\mathbb{E}}_{q_\eta}[T(x)T(x)^{\mathrm{T}}] - \mathop{\mathbb{E}}_{q_\eta}[T(x)] \mathop{\mathbb{E}}_{q_\eta}[T(x)]^{\mathrm{T}}.\end{aligned} \tag{24}$$

It goes by the definition of covariance that the Hessian matrix $\nabla_\eta^2 A$ is positive semi-definite and therefore $A(\cdot)$ is a convex function. □

Corollary 1. *The natural parameter space Ω is a convex set.*

Proposition 2. *The gradient of the log partition function $\nabla A : \Omega \to \mathcal{M}$ is a one-to-one mapping if and only if the exponential representation is minimal.*

Proof. For a proof see [25, Proposition 3.2]. □

Proposition 3. *In a minimal exponential family the gradient map ∇A is onto a the interior of $\mathcal{M}$ denoted by $\mathcal{M}^\circ$. Consequently, for each $\mu \in \mathcal{M}^\circ$, there exists some $\eta = \eta(\mu) \in \Omega$ such that $\mathbb{E}_{q_\eta}[T(x)] = \mu$.*

Proof. For a proof see [25, Theorem 3.3]. □

A. Merging algorithm for the exponential family

In Theorem 3 we derive an expression for approximating a mixture consisting of a member of the exponential family by a single component of the same member of the exponential family.

Theorem 3. *For any finite mixture density $p_{\mathcal{L}}(x)$ as in* (3) *where its basic densities $q(x;\eta)$ belong to regular exponential family of distributions and have minimal representation there exists a unique permissible natural parameter $\eta^* \in \Omega$ minimizing the Kullback-Leibler divergence $D_{KL}(p_{\mathcal{L}}||q_L)$ given by solving the system of equations*

$$\nabla_{\eta^L} A(\eta^L) = \sum_{I\in\mathcal{L}} \widehat{w}^I \nabla_{\eta^I} A(\eta^I). \tag{25}$$

Proof. To perform the minimization of the KLD we write down the necessary conditions for optimality also known as Karush-Kuhn-Tucker (KKT) conditions for the unconstrained problem

$$\nabla_{\eta^L} D_{KL}(p_{\mathcal{L}}||q_L) = 0, \tag{26}$$

and will show that the solution for the unconstrained problem is a permissible solution to the constrained problem i.e., $\eta^* \in \Omega$. Now we will simplify the gradient of the KLD given in equation (26) in

$$\begin{aligned}
\nabla_{\eta^L} D_{KL}(p_{\mathcal{L}}||q_L) &= \nabla_{\eta^L} \mathop{\mathbb{E}}_{p_{\mathcal{L}}}[\log \frac{p_{\mathcal{L}}}{q_L}] \\
&= -\nabla_{\eta^L} \mathop{\mathbb{E}}_{p_{\mathcal{L}}}[\log q_L] \\
&= -\nabla_{\eta^L} \mathop{\mathbb{E}}_{p_{\mathcal{L}}}[\log h(x) + \eta^L \cdot T(x) - A(\eta^L)] \\
&= -\mathop{\mathbb{E}}_{p_{\mathcal{L}}}[T(x) - \nabla_{\eta^L} A(\eta^L)] = -\mathop{\mathbb{E}}_{p_{\mathcal{L}}}[T(x)] + \nabla_{\eta^L} A(\eta^L) \\
&= -\sum_{I\in\mathcal{L}} \widehat{w}^I \nabla_{\eta^I} A(\eta^I) + \nabla_{\eta^L} A(\eta^L).
\end{aligned} \tag{27}$$

Therefore the KKT condition of equation (26) is equivalent to

$$\nabla_{\eta^L} A(\eta^L) = \sum_{I\in L} \widehat{w}^I \nabla_{\eta^I} A(\eta^I). \tag{28}$$

Since $\nabla^2_{\eta^L} D_{KL}(p_{\mathcal{L}}||q_L) = \nabla^2_{\eta^L} A(\eta^L)$ and $\nabla^2_{\eta^L} A(\eta^L)$ is positive definite according to Proposition 1 and Ω is an open and convex set the optimization problem is a convex optimization problem and KKT necessary conditions for optimality are sufficient conditions for optimality and the solution is unique. Now it remains to show that η^L is a permissible solution to the constrained problem which follows from Proposition 3. □

Corollary 2. *The merging algorithm of equation* (25) *is equivalent to matching the expectations of sufficient statistics $T(x)$ with respect to the two densities $p_{\mathcal{L}}$ and q_L i.e.,*

$$\mathop{\mathbb{E}}_{q_L}[T(x)] = \mathop{\mathbb{E}}_{p_{\mathcal{L}}}[T(x)]. \tag{29}$$

The approximation of a density with a density belonging to the exponential family is derived in [17, Section 10.7] in the context of expectation propagation. But, the feasibility of the solution to the minimization problem and the fact that the solution is a global minimum is not discussed. In [26], the convexity of the cost function is proven but reference to the minimality of representation of the exponential family which is needed for the uniqueness of the solution and elaboration on the feasibility of the solution are missing. In [25], all these aspects are discussed for approximating a general density with a member of the exponential family. Here we have used the general results of [25] for the specific problem of mixture reduction.

In the following example, the expressions needed for the nonlinear system of equations which arise in reduction of mixtures of Gaussian inverse Wishart distribution, which is used in the extended target tracking framework [27] are given.

Example 3: Consider the Gaussian inverse Wishart density

$$\mathrm{GIW}(x, X; m, P, \nu, \Psi) = \mathcal{N}(x; m, P)\, \mathcal{IW}_d(X; \nu, \Psi) \tag{30}$$

where $m \in \mathbb{R}^k$ denotes the mean value, $P \in S^k_+$ is the covariance matrix, $\nu > 2d$ is the degrees of freedom and $\Psi \in S^d_{++}$ is the scale matrix. For the GIW distribution we have

$$\eta = \left(-\frac{1}{2}\nu, -\frac{1}{2}\Psi, P^{-1}m, -\frac{1}{2}P^{-1}\right) \tag{31}$$

and

$$A(\eta) = \left(\eta_1 + \frac{d+1}{2}\right)\log|-\eta_2| + \log\Gamma_d\left(-\eta_1 - \frac{d+1}{2}\right) - \frac{1}{4}\eta_3^{\mathrm{T}}\eta_4^{-1}\eta_3 - \frac{1}{2}\log|-2\eta_4|, \tag{32}$$

where $\Gamma_d(\cdot)$ is the multivariate gamma function. Notice that the natural parameter is not a vector since its elements can be matrix valued. The gradient $\nabla_\eta A(\eta)$, should be substituted in (25) to solve for the parameters of the merged component and is given by

$$\begin{aligned}\nabla_\eta A(\eta) &= \left(\log|-\eta_2| - \psi_d\left(-\eta_1 - \frac{d+1}{2}\right), \left(\eta_1 + \frac{d+1}{2}\right)\eta_2^{-1}, -\frac{1}{2}\eta_3^{\mathrm{T}}\eta_4^{-1}, \frac{1}{4}\eta_4^{-\mathrm{T}}\eta_3\eta_3^{\mathrm{T}}\eta_4^{-\mathrm{T}} - \frac{1}{2}\eta_4^{-1}\right) \\ &= \left(\log|\frac{1}{2}\Psi| - \psi_d\left(\frac{\nu-d-1}{2}\right), (\nu-d-1)\Psi^{-1}, m^{\mathrm{T}}, mm^{\mathrm{T}} + P\right) \end{aligned} \tag{33}$$

where $\psi_d(\cdot)$ is the multivariate digamma function. Also the expression for $\mathcal{B}(I,J)$ for the GIW density is given as

$$\mathcal{B}_{\mathrm{GIW}}(I,J) = w^{IJ}\phi(\nu^{IJ}, P^{IJ}) - w^I\phi(\nu^I, P^I) - w^J\phi(\nu^J, P^J), \tag{34}$$

where,

$$\phi(\nu, P) = -\frac{\nu-d-1}{2}\psi_d\left(\frac{\nu-d-1}{2}\right) + \log\Gamma_d\left(\frac{\nu-d-1}{2}\right) + \frac{d}{2}\nu + \frac{1}{2}\log|P|.$$

The expression for the $Q(I,J)$ which is needed for the calculation of the ISE between two mixtures can be simplified to

$$Q(I,J) = \exp\left(A(\eta^I + \eta^J) - A(\eta^I) - A(\eta^J)\right) \tag{35}$$

since $\mathbb{E}_{q(x;\eta^I+\eta^J)}[h(x)] = 1$.

The expression for SKL for the GIW density can be obtained as [18]

$$D_{SKL}(I,J) = (\eta^I - \eta^J)\cdot(\nabla_{\eta^I}A(\eta^I) - \nabla_{\eta^J}A(\eta^J)). \tag{36}$$

■

B. Numerical Simulations

Three mixture reduction algorithms (MRAs), AKL, SKL and ISE [18], are compared in numerical simulation for mixtures of the most common members of the exponential family distributions. A total of 1000 random mixture densities are randomly generated for each density and they are reduced with different reduction aggressiveness, namely, reduction from 25 components down to $\mathbf{M}$ components where $25 > \mathbf{M} > 1$. The reduced mixtures are compared to the original mixtures in terms of ISE between the reduced mixture and the original mixture, calculated analytically. Due to the high variance of the ISE values, which in turn is due to the the high variability in random mixture density parameters used for the MC simulation, the comparison population mean and standard deviation does not show the difference between the MRAs. As a remedy, a paired difference test is used.

The Wilcoxon signed-rank test [28] is a non-parametric paired difference test which can be used for comparing two matched samples. The Wilcoxon signed-rank test statistics is affected by both the magnitude and the sign of the difference between the matched samples. It can be used when a population cannot be assumed to be normally distributed which is required for paired Student's t-test [28]. When the paired difference is statistically significant (p-value $< 1\%$) the null hypothesis (the paired ISE values come from continuous distributions with equal medians) is rejected. The Wilcoxon signed-rank test is implemented in this numerical simulation using the *ranksum* command in MATLAB®.

In Tables III to XVI the decimal logarithm of the p-value for the two sided Wilcoxon rank sum test is given for ISE values of all pairs of MRAs i.e., AKL versus SKL, AKL versus ISE and SKL versus ISE. When the ISE values corresponding to two MRAs have a significant difference, the MRA whose ISE value has a smaller median is pointed out. The latter comparison is performed using the one-sided Wilcoxon rank sum test.

Similar comparison is performed for all pairs of MRAs where the error between the original and the approximate mixture is the KLD between the original mixture and the approximate mixture. Due to the absence of an analytical solution, the KLD is calculated numerically via MC sampling method. The number of samples used in the numerical calculation of the KLD is given for each type of density in the corresponding section. In the following using an example the simulation scenario will be further explained.

Consider the simulation parameters and results given in section III-B1 for exponential distribution. 1000 random mixtures of exponential distribution are generated. The parameters of the MC simulation are given in (37). For all 1000 MC samples the approximation of the original density with 25 components with a mixture consisting of $25 > \mathbf{M} > 1$ components using AKL, SKL and ISE is calculated. For each MC realization and $\mathbf{M}$, the integral square error as well as the KLD between the original mixture and the approximate mixture is calculated. The objective of the simulations is to find out, firstly, whether the two MRAs in each pair are different (for the simulation parameter range) or not. Secondly, when they are different which MRA gives lower error which in our case is the KLD and the ISE between the original mixture and its approximation. The p-value for the two sided Wilcoxon rank sum test for each pair of MRAs is used to determine whether the two MRAs are different in terms of the error. The decimal logarithm of the p-values are given in Table III. When the difference between the errors of the two MRAs in a pair is statistically significant it is relevant to compare the median of the errors of the two MRAs in a pair to find out which one is smaller. Using the one-sided Wilcoxon signed-rank test, where the alternative hypothesis states that the median error of one MRA in a pair is less than the median error of the other MRA in the same pair the MRA with smaller error is identified and pointed out in Table III. In the first three columns the pairwise comparison is performed with

respect to the KLD as a measure of error and in the last three columns the pairwise comparison is performed with respect to the integral square error. For example, the first column of Table III shows that the difference between AKL and SKL is not significant for reductions from 25 components down to 13 components while AKL is significantly better than SKL for more aggressive reductions such as 25 down to $\mathbf{M}$ where $13 > \mathbf{M} > 1$. In Tables III to XVI the winner MRA is indicated using its respective symbols which is given in the following. AKL is denoted by ★, SKL is denoted by ♦ and ISE is denoted by ■.

A similar comparison is made for 13 other mixture densities in Sections III-B2 to III-B14. The parameters of the Monte-Carlo simulations are given for each type of density in the corresponding subsection.

1) Exponential Distribution: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \operatorname{Exp}(x; \lambda_j^I), \text{ where,} \tag{37a}$$

$$\operatorname{Exp}(x; \lambda) = \lambda \exp(-\lambda x), \tag{37b}$$

$$N = 25, \tag{37c}$$

$$\frac{1}{\lambda_j^I} \sim \mathcal{U}(0.5, 50.5), \tag{37d}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I \right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{37e}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 3.

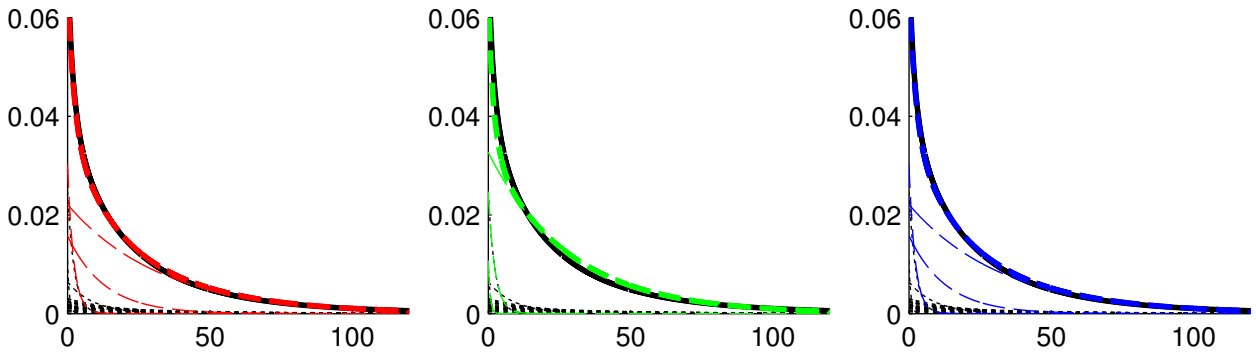


Figure 3. Mixture of Exponential Distribution: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm.

The pairwise comparison of the MRAs in the MC simulation is given in Table III. In Table III, the difference between MRAs becomes more significant as the number of components in the approximate density decreases. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL. AKL has the best overall performance, wheras SKL has the worst performance in all four pairwise comparisons.

The average cycle times for MRAs in the MC simulations are given in Figure 4. The ISE is the most costly MRA and SKL is the least costly MRA.

Table III
EXPONENTIAL DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.0245 –	-0.1938 –	-0.1598 –	-0.2421 –	-0.0312 –	-0.2893 –
23	-0.2288 –	-0.0018 –	-0.2201 –	-0.7464 –	-0.1712 –	-1.1112 –
22	-0.1754 –	-0.0829 –	-0.2659 –	-1.1143 –	-0.2081 –	-1.6396 –
21	-0.2539 –	-0.1221 –	-0.1034 –	-1.3173 –	-0.2379 –	-1.9472 –
20	-0.0618 –	-0.2490 –	-0.1552 –	-2.4956 ★	-0.2157 –	-3.2786 ■
19	-0.1081 –	-0.0747 –	-0.0251 –	-2.6226 ★	-0.2833 –	-3.6424 ■
18	-0.0212 –	-0.0024 –	-0.0144 –	-3.3577 ★	-0.3151 –	-4.6512 ■
17	-0.0765 –	-0.0582 –	-0.1419 –	-3.5792 ★	-0.4558 –	-5.4388 ■
16	-0.0793 –	-0.0965 –	-0.0094 –	-3.5077 ★	-0.5350 –	-5.5530 ■
15	-0.1931 –	-0.2410 –	-0.0259 –	-2.7408 ★	-0.6561 –	-4.9214 ■
14	-0.5629 –	-0.1484 –	-0.3069 –	-3.2014 ★	-0.5657 –	-5.2842 ■
13	-1.1858 –	-0.5753 –	-0.3421 –	-2.7337 ★	-0.6885 –	-5.0495 ■
12	-2.4986 ★	-0.8007 –	-0.8842 –	-2.4614 ★	-0.8116 –	-4.9965 ■
11	-4.8916 ★	-1.6517 –	-1.4433 –	-2.7761 ★	-0.8506 –	-5.5277 ■
10	-9.5861 ★	-4.4022 ★	-1.5928 –	-2.5546 ★	-0.9246 –	-5.5040 ■
9	-18.1821 ★	-5.9043 ★	-4.3421 ■	-2.0819 ★	-1.3032 –	-5.6117 ■
8	-22.8025 ★	-6.7998 ★	-5.6786 ■	-2.1383 ★	-1.2285 –	-5.5299 ■
7	-24.2039 ★	-6.3701 ★	-6.8158 ■	-1.4000 –	-1.1681 –	-4.2005 ■
6	-34.4967 ★	-9.0905 ★	-9.7050 ■	-1.0345 –	-1.6989 –	-4.4325 ■
5	-32.6684 ★	-6.6254 ★	-11.6244 ■	-1.1463 –	-1.6513 –	-4.5305 ■
4	-37.2092 ★	-7.2389 ★	-12.5136 ■	-1.1834 –	-1.9360 –	-5.3096 ■
3	-38.7857 ★	-4.7452 ★	-18.8732 ■	-3.0600 ★	-1.1579 –	-6.8602 ■
2	-35.7668 ★	-3.3114 ★	-18.8536 ■	-2.6730 ★	-1.1859 –	-6.4104 ■

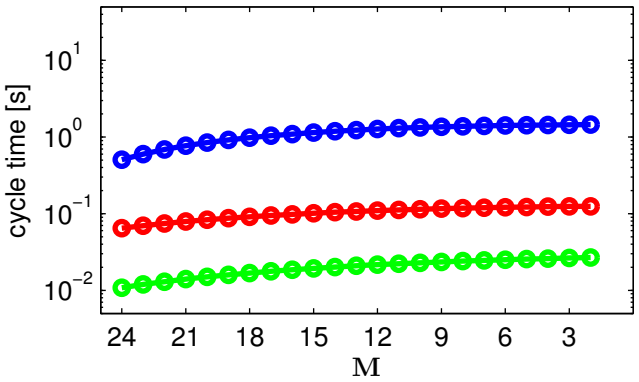


Figure 4. Exponential Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

2) Weibull Distribution with known shape k: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \,\text{Weibull}(x; \lambda_j^I, k_j), \text{ where,} \tag{38a}$$

$$\text{Weibull}(x; \lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp(-\frac{x^k}{\lambda^k}), \tag{38b}$$

$$k_j \sim \mathcal{U}(1, 11) \tag{38c}$$

$$N = 25, \tag{38d}$$

$$\lambda_j^I \sim \mathcal{U}(0.1, 50.1), \tag{38e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{38f}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^S \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 5.

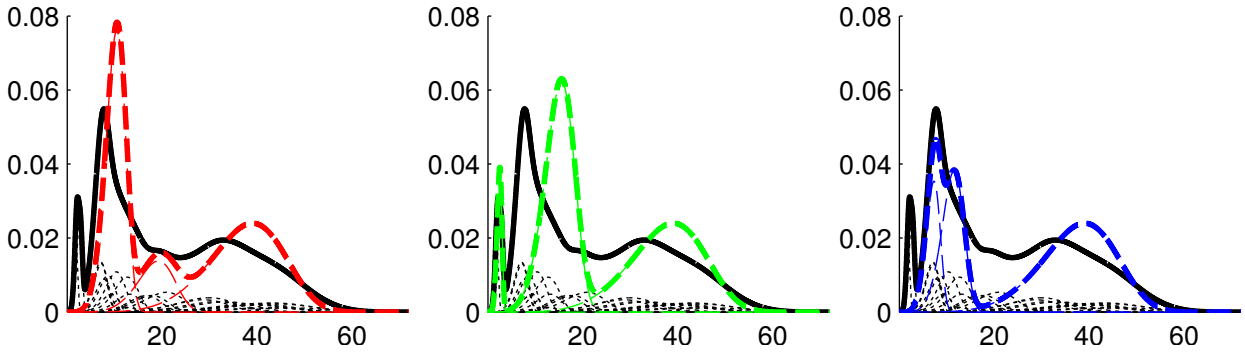


Figure 5. Mixture of Weibull Distribution with known shape k: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm.

The pairwise comparison of the MRAs in the MC simulation is given in Table IV. In Table IV, the difference between MRAs becomes more significant as the number of components in the approximate density decreases. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL. AKL obtains lower ISE compared to SKL which confirms that AKL is better MRA compared to SKL for Weibull distribution with known shape. When SKL and ISE are compared with respect to KLD between the original mixture and its approximation, as the reduction aggressiveness increases, the MRA with lower error changes to SKL from ISE. This phenomenon can be attributed to the property ISE has and does not exist in SKL. In the ISE MRA the mixture can be reduced via pruning, whereas in the other two algorithms pruning does not exist. Since the pruning (especially in a multimodal distribution) can significantly remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, the SKL can obtain smaller KLD between the original mixture and its approximation.

Table IV
WEIBULL DISTRIBUTION WITH KNOWN SHAPE k: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.0001 –	-0.0090 –	-0.0093 –	-0.1911 –	-0.0324 –	-0.2358 –
23	-0.0641 –	-0.0528 –	-0.0112 –	-0.4126 –	-0.0862 –	-0.5625 –
22	-0.0747 –	-0.2245 –	-0.1224 –	-0.6492 –	-0.2181 –	-1.0729 –
21	-0.2844 –	-0.0547 –	-0.2028 –	-0.7501 –	-0.2766 –	-1.3150 –
20	-0.6152 –	-0.4015 –	-0.1470 –	-1.2737 –	-0.2730 –	-1.9845 –
19	-1.5012 –	-0.0755 –	-1.2792 –	-1.4833 –	-0.3644 –	-2.4514 ■
18	-1.2408 –	-0.1939 –	-0.8337 –	-1.6070 –	-0.4440 –	-2.8146 ■
17	-1.7868 –	-0.3668 –	-0.9723 –	-1.6660 –	-0.5759 –	-3.2161 ■
16	-2.3510 ★	-0.4892 –	-1.2665 –	-1.6830 –	-0.6268 –	-3.3351 ■
15	-2.3091 ★	-0.1783 –	-1.7569 –	-1.7823 –	-0.5735 –	-3.3776 ■
14	-2.0010 ★	-0.2371 –	-1.3750 –	-1.6679 –	-0.7693 –	-3.6541 ■
13	-2.4272 ★	-0.2173 –	-1.7742 –	-1.6398 –	-0.9620 –	-3.9383 ■
12	-3.6589 ★	-0.2721 –	-2.6384 ■	-1.8456 –	-1.1062 –	-4.6680 ■
11	-3.1574 ★	-0.3915 –	-1.9271 –	-1.4695 –	-1.1391 –	-4.0800 ■
10	-3.6981 ★	-0.4811 –	-2.1403 ■	-1.9347 –	-1.2835 –	-5.0476 ■
9	-3.7709 ★	-0.7694 –	-1.5766 –	-2.0866 ★	-1.6311 –	-6.0823 ■
8	-4.2196 ★	-1.2522 –	-1.1592 –	-2.7513 ★	-1.7477 –	-7.7040 ■
7	-5.0170 ★	-2.0478 ★	-0.8010 –	-3.3091 ★	-2.8638 ■	-10.5665 ■
6	-6.3224 ★	-3.7271 ★	-0.3622 –	-4.9958 ★	-5.1138 ■	-17.2348 ■
5	-7.2148 ★	-6.6388 ★	-0.2352 –	-6.4480 ★	-8.5747 ■	-24.7890 ■
4	-8.3432 ★	-11.5138 ★	-1.5494 –	-8.7835 ★	-16.5768 ■	-41.7100 ■
3	-11.6261 ★	-20.0118 ★	-3.5781 ♦	-12.2249 ★	-30.2593 ■	-69.4317 ■
2	-10.0793 ★	-32.6085 ★	-12.7112 ♦	-8.2912 ★	-56.7129 ■	-97.0374 ■

The average cycle times for MRAs in the MC simulations are given in Figure 6. The ISE is the most costly MRA and SKL is the least costly MRA.

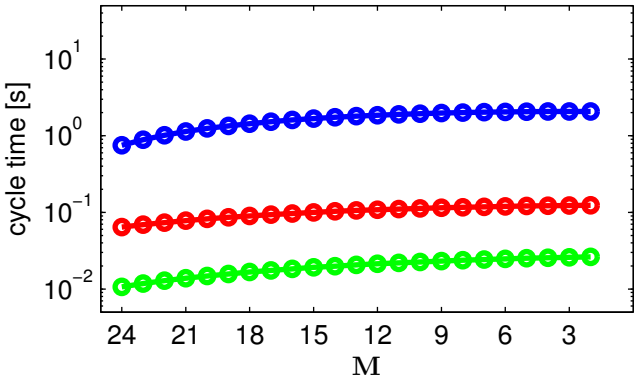


Figure 6. Weibull Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

3) Laplace Distribution with known mean μ: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \,\text{Laplace}(x;\mu,b_j^I), \text{ where,} \tag{39a}$$

$$\text{Laplace}(x;\mu,b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right), \tag{39b}$$

$$\mu = 5, \tag{39c}$$

$$N = 25, \tag{39d}$$

$$b_j^I \sim \mathcal{U}(0.5, 50.5), \tag{39e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{39f}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 7.

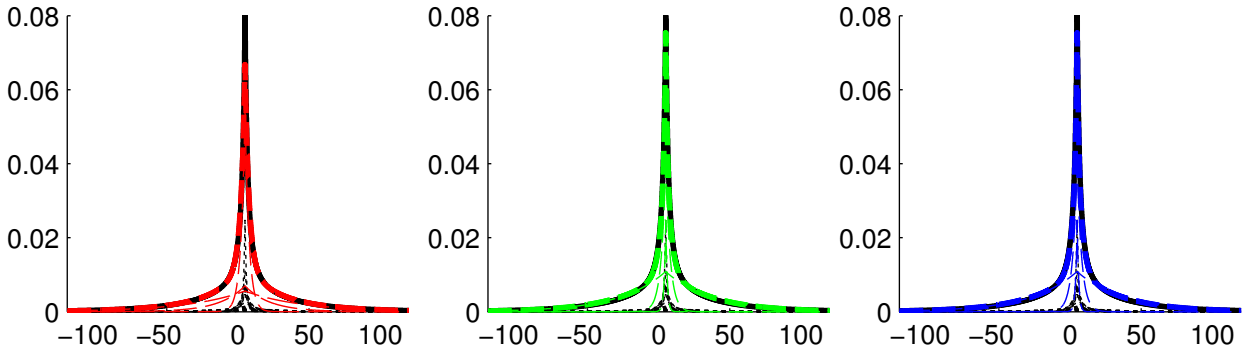


Figure 7. Mixture of Laplace Distribution with known mean μ: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm.

The pairwise comparison of the MRAs in the MC simulation is given in Table V. The simulation result is similar to the exponential distribution as expected; In Table V, the difference between MRAs becomes more significant as the number of components in the approximate density decreases. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it

Table V

LAPLACE DISTRIBUTION WITH KNOWN MEAN μ: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.0962 –	-0.0650 –	-0.0263 –	-0.2372 –	-0.0179 –	-0.2650 –
23	-0.1997 –	-0.0964 –	-0.0836 –	-0.7021 –	-0.0820 –	-0.8829 –
22	-0.0586 –	-0.0564 –	-0.0063 –	-1.1175 –	-0.1893 –	-1.5922 –
21	-0.1331 –	-0.1300 –	-0.0022 –	-1.4195 –	-0.1812 –	-1.9244 –
20	-0.1152 –	-0.1518 –	-0.2795 –	-1.7599 –	-0.2628 –	-2.5433 ■
19	-0.0151 –	-0.0388 –	-0.0175 –	-2.3642 ★	-0.3088 –	-3.4407 ■
18	-0.0015 –	-0.2135 –	-0.1802 –	-3.0355 ★	-0.4348 –	-4.6073 ■
17	-0.0173 –	-0.1325 –	-0.1439 –	-2.8340 ★	-0.4797 –	-4.4775 ■
16	-0.0338 –	-0.0247 –	-0.0185 –	-3.4486 ★	-0.4746 –	-5.2695 ■
15	-0.1529 –	-0.1606 –	-0.0038 –	-4.0804 ★	-0.4296 –	-5.9451 ■
14	-0.7908 –	-0.5748 –	-0.1168 –	-4.0948 ★	-0.5483 –	-6.3319 ■
13	-1.1292 –	-0.6289 –	-0.2622 –	-3.1211 ★	-0.7468 –	-5.7828 ■
12	-2.9637 ★	-1.2281 –	-0.8113 –	-3.1443 ★	-0.9489 –	-6.3642 ■
11	-5.1150 ★	-0.9568 –	-2.3108 ■	-3.0705 ★	-0.7493 –	-5.6555 ■
10	-9.2361 ★	-2.5652 ★	-2.8147 ■	-2.6958 ★	-0.9498 –	-5.7725 ■
9	-15.6072 ★	-4.4253 ★	-4.3395 ■	-1.9665 –	-1.2768 –	-5.3992 ■
8	-25.0252 ★	-5.7353 ★	-7.9421 ■	-1.5576 –	-1.4513 –	-4.9538 ■
7	-29.8030 ★	-6.7014 ★	-9.4515 ■	-1.8908 –	-1.4373 –	-5.6010 ■
6	-27.9109 ★	-8.1227 ★	-6.8745 ■	-1.0099 –	-1.6142 –	-4.2521 ■
5	-32.0328 ★	-7.4151 ★	-9.5161 ■	-0.9609 –	-1.6796 –	-4.3043 ■
4	-31.0722 ★	-5.3634 ★	-12.5938 ■	-1.4256 –	-1.6284 –	-5.1172 ■
3	-40.6181 ★	-5.1522 ★	-18.2919 ■	-1.2883 –	-1.9317 –	-5.4725 ■
2	-31.6808 ★	-3.3114 ★	-15.8921 ■	-2.8625 ★	-0.8027 –	-5.6460 ■

obtains lower KLD compared to ISE and SKL. AKL has the best overall performance, whereas SKL has the worst performance in all four pairwise comparisons it is being compared.

The average cycle times for MRAs in the MC simulations are given in Figure 8. The ISE is the most costly MRA and SKL is the least costly MRA.

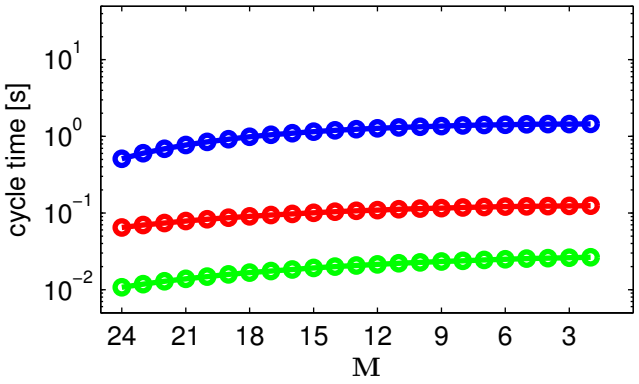


Figure 8. Laplace Distribution with known mean μ: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

4) Rayleigh Distribution : For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \,\text{Rayleigh}(x;\sigma_j^I), \text{ where,} \tag{40a}$$

$$\text{Rayleigh}(x;\sigma) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{40b}$$

$$N = 25, \tag{40c}$$

$$\sigma_j^I \sim \mathcal{U}(0.5, 10.5), \tag{40d}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{40e}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 9.

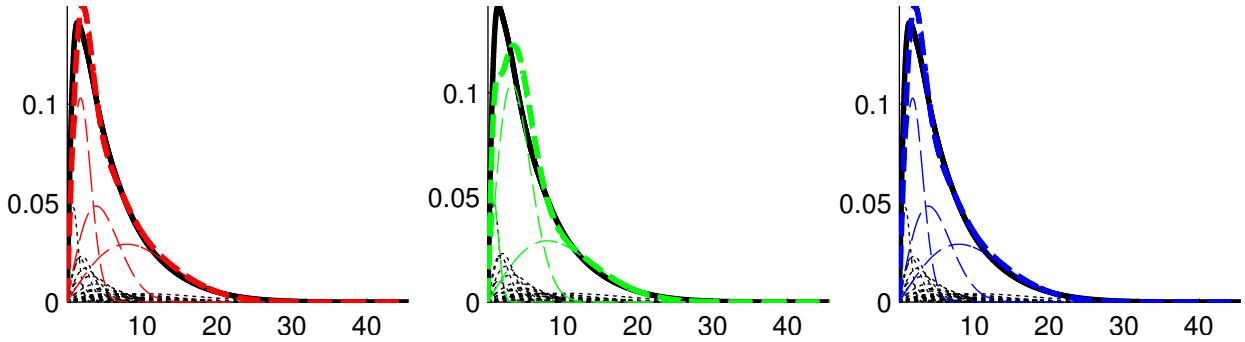


Figure 9. Mixture of Rayleigh Distribution: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm.

The pairwise comparison of the MRAs in the MC simulation is given in Table VI. In Table VI, the difference between MRAs becomes more significant as the number of components in the approximate density decreases. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL. AKL has the best overall performance, whereas SKL has the worst performance in all four pairwise comparisons it is being compared.

The average cycle times for MRAs in the MC simulations are given in Figure 10. The ISE is the most costly MRA and SKL is the least costly MRA.

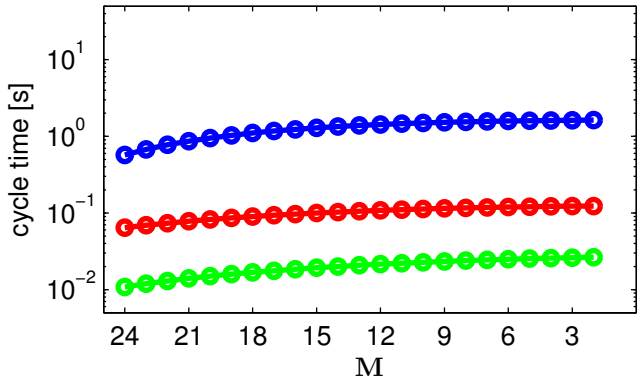


Figure 10. Rayleigh Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

Table VI
RAYLEIGH DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			comparison with respect to ISE		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.1626 –	-0.1794 –	-0.0124 –	-0.1774 –	-0.0352 –	-0.2258 –
23	-0.0010 –	-0.0538 –	-0.0562 –	-0.5656 –	-0.1577 –	-0.8712 –
22	-0.0768 –	-0.2913 –	-0.3825 –	-1.2550 –	-0.2280 –	-1.8394 –
21	-0.0964 –	-0.0770 –	-0.1835 –	-1.4594 –	-0.2569 –	-2.1556 ■
20	-0.2526 –	-0.0286 –	-0.2859 –	-2.1461 ★	-0.4131 –	-3.4445 ■
19	-0.2112 –	-0.1637 –	-0.0351 –	-3.7503 ★	-0.4156 –	-5.4295 ■
18	-0.1636 –	-0.1713 –	-0.0038 –	-4.9468 ★	-0.5570 –	-7.3597 ■
17	-0.2893 –	-0.3353 –	-0.0423 –	-5.5607 ★	-0.5425 –	-8.1421 ■
16	-1.8478 –	-0.7208 –	-0.5971 –	-5.5612 ★	-0.7415 –	-8.8638 ■
15	-1.7299 –	-0.9307 –	-0.3843 –	-5.7804 ★	-0.8536 –	-9.4606 ■
14	-2.7756 ★	-0.6957 –	-1.2247 –	-7.1727 ★	-0.8813 –	-11.4796 ■
13	-6.4301 ★	-1.6853 –	-2.0795 ■	-6.4333 ★	-1.2790 –	-11.7961 ■
12	-10.1006 ★	-3.7478 ★	-2.2290 ■	-6.8267 ★	-1.2241 –	-12.0098 ■
11	-14.7225 ★	-2.4877 ★	-5.9618 ■	-8.7181 ★	-0.9841 –	-13.8364 ■
10	-17.3903 ★	-3.0891 ★	-6.7822 ■	-7.8278 ★	-1.3538 –	-14.0045 ■
9	-25.2586 ★	-3.9180 ★	-10.3622 ■	-8.6619 ★	-1.3713 –	-15.4868 ■
8	-28.4945 ★	-6.4986 ★	-8.8325 ■	-8.6525 ★	-1.7343 –	-16.4593 ■
7	-30.8952 ★	-4.2157 ★	-14.3242 ■	-9.0076 ★	-1.6616 –	-16.7751 ■
6	-30.9283 ★	-3.0021 ★	-16.3437 ■	-8.1166 ★	-1.7301 –	-15.9721 ■
5	-32.6869 ★	-2.7005 ★	-18.1820 ■	-8.0696 ★	-1.3293 –	-14.9522 ■
4	-26.8932 ★	-2.0226 ★	-15.9112 ■	-6.6795 ★	-1.1010 –	-12.2188 ■
3	-34.2745 ★	-1.4270 –	-23.4657 ■	-10.6730 ★	-1.1537 –	-17.0686 ■
2	-29.4471 ★	-0.9209 –	-20.7546 ■	-17.5143 ★	-0.2132 –	-20.5881 ■

5) Log-normal Distribution: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \log -\mathcal{N}(x; \mu_j^I, \sigma_j^I), \text{ where,} \tag{41a}$$

$$\log -\mathcal{N}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\log x - \mu)^2\right), \tag{41b}$$

$$N = 25, \tag{41c}$$

$$\mu_j^I \sim \mathcal{U}(0, 10), \tag{41d}$$

$$(\sigma_j^I)^2 \sim \mathcal{U}(10^{-5}, 10^{-1}), \tag{41e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{41f}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 11.

The pairwise comparison of the MRAs in the MC simulation is given in Table VII. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE. When SKL and ISE are compared with respect to KLD between the original mixture and its approximation the SKL obtains lower KLD value which can be attributed to the property ISE has and does not exist in SKL. In the ISE the mixture can be reduced via pruning, whereas in the other two algorithms pruning does not exist. Since the pruning can remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, the SKL can obtain smaller KLD between the original mixture and its approximation. Also, the log-normal mixture densities of this simulation have large spread over the support compared to other densities which will further amplify the sensitivity to pruning. Although AKL seems to have a better overall performance compared to SKL with respect to KLD, the best MRA switches back and forth for different reduction aggressiveness which undermines the quantitative evaluation.

The average cycle times for MRAs in the MC simulations are given in Figure 12. The ISE is the most costly MRA and SKL is the least costly MRA.

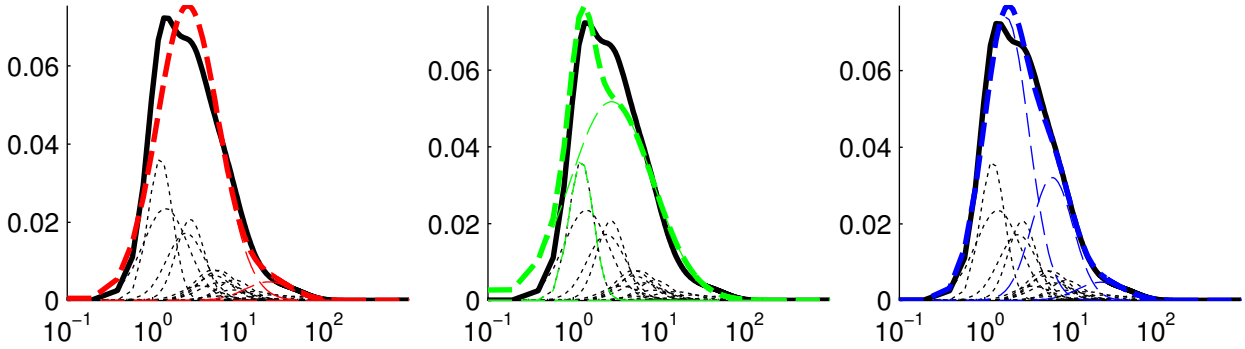


Figure 11. Mixture of Log-normal Distribution: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm. The x-axis is in the logarithmic scale to enhance the illustration.

Table VII
LOG-NORMAL DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.6145 –	-1.8341 –	-0.7144 –	-0.3846 –	-48.0554 ■	-52.6655 ■
23	-1.2480 –	-7.4887 ★	-3.8385 ♦	-1.0664 –	-106.1949 ■	-117.0927 ■
22	-2.3871 ★	-12.8058 ★	-5.9990 ♦	-1.2884 –	-170.8586 ■	-181.3566 ■
21	-3.6242 ★	-18.8469 ★	-8.6582 ♦	-2.1512 ★	-196.1066 ■	-206.9163 ■
20	-4.9855 ★	-23.0318 ★	-9.6055 ♦	-2.1449 ★	-225.1425 ■	-231.1083 ■
19	-10.1000 ★	-33.4158 ★	-11.9344 ♦	-2.5868 ★	-241.8211 ■	-251.9255 ■
18	-8.3365 ★	-39.6197 ★	-17.5483 ♦	-3.2056 ★	-256.5586 ■	-266.4581 ■
17	-8.8301 ★	-49.7039 ★	-24.8908 ♦	-2.7108 ★	-262.3280 ■	-273.6369 ■
16	-7.2943 ★	-57.0799 ★	-32.6845 ♦	-2.7288 ★	-271.4055 ■	-277.0616 ■
15	-10.0311 ★	-64.8991 ★	-35.6211 ♦	-3.0195 ★	-275.1351 ■	-278.4623 ■
14	-11.5418 ★	-75.5506 ★	-42.5355 ♦	-3.4414 ★	-279.8285 ■	-282.7999 ■
13	-7.8877 ★	-79.4361 ★	-52.5541 ♦	-2.7176 ★	-279.6990 ■	-282.9228 ■
12	-4.8917 ★	-89.7251 ★	-68.5670 ♦	-0.7825 –	-280.5910 ■	-274.5060 ■
11	-2.6310 ★	-95.6595 ★	-80.0980 ♦	-1.1485 –	-278.9853 ■	-266.3671 ■
10	-1.3197 –	-116.0848 ★	-107.3736 ♦	-0.3856 –	-272.4362 ■	-263.6656 ■
9	-0.0612 –	-141.1819 ★	-141.9355 ♦	-0.2965 –	-262.1799 ■	-249.7839 ■
8	-0.9060 –	-174.3024 ★	-181.0961 ♦	-0.2002 –	-251.6548 ■	-233.1286 ■
7	-3.8672 ♦	-184.4788 ★	-198.6944 ♦	-0.4324 –	-242.4126 ■	-224.8731 ■
6	-1.6439 –	-207.6271 ★	-215.4164 ♦	-0.0984 –	-225.6450 ■	-212.8376 ■
5	-0.0173 –	-238.3513 ★	-239.7716 ♦	-0.1921 –	-210.3870 ■	-195.4918 ■
4	-0.1303 –	-249.5171 ★	-250.7162 ♦	-0.1909 –	-216.5991 ■	-195.0157 ■
3	-0.1598 –	-258.9098 ★	-257.0986 ♦	-0.3233 –	-199.0075 ■	-184.6611 ■
2	-6.5654 ★	-276.4668 ★	-270.1903 ♦	-1.0341 –	-184.6181 ■	-172.3660 ■

6) Gamma Distribution: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \,\mathrm{Gamma}(x; \alpha_j^I, \beta_j^I), \text{ where,} \tag{42a}$$

$$\mathrm{Gamma}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \tag{42b}$$

$$N = 25, \tag{42c}$$

$$\alpha_j^I \sim \mathcal{U}(10, 60), \tag{42d}$$

$$\beta_j^I \sim \mathcal{U}(1, 11), \tag{42e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I \right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{42f}$$

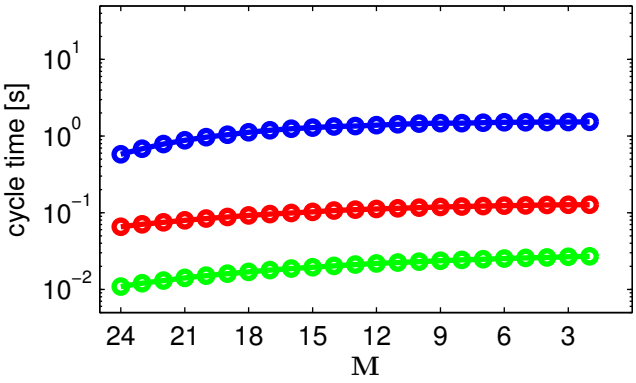


Figure 12. Log-normal Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 13.

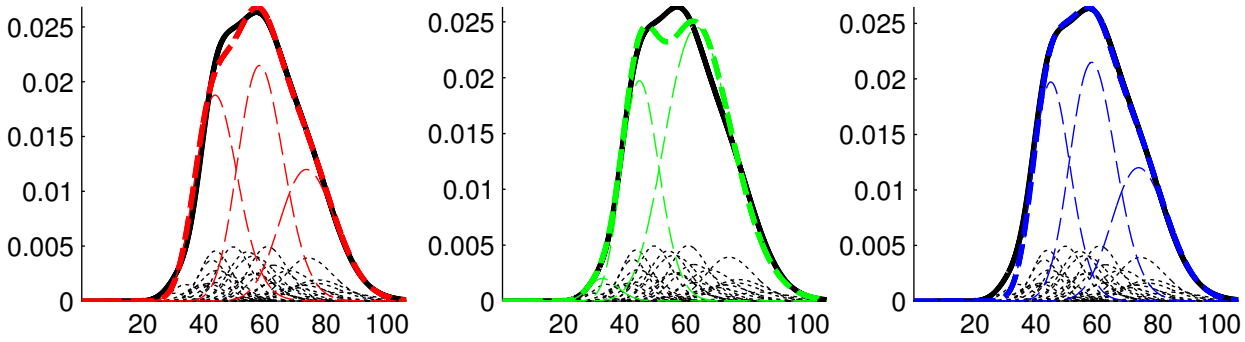


Figure 13. Mixture of Gamma Distribution: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm.

The pairwise comparison of the MRAs in the MC simulation is given in Table VIII. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. AKL outperforms SKL in terms of ISE. All three MRAs are nearly equal with respect to KLD. The superior performance of SKL and AKL compared to ISE at very aggressive reductions can be again attributed to the pruning in the ISE approach. Although ISE seems to have a better overall performance compared to SKL and AKL with respect to KLD, the best MRA for a given application scenario should be selected based on the reduction aggressiveness used in the application.

The average cycle times for MRAs in the MC simulations are given in Figure 14. The ISE is the most costly MRA and SKL is the least costly MRA.

Table VIII
GAMMA DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			comparison with respect to ISE		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.0116 –	-1.8503 –	-1.6496 –	-1.5323 –	-40.0138 ■	-48.6233 ■
23	-0.4329 –	-6.9135 ■	-8.5306 ■	-5.5893 ★	-95.9699 ■	-125.4864 ■
22	-3.2114 ★	-9.3560 ■	-21.8774 ■	-10.6424 ★	-151.9538 ■	-198.2572 ■
21	-4.6637 ★	-24.7910 ■	-45.5473 ■	-12.8846 ★	-203.9425 ■	-253.1017 ■
20	-5.9074 ★	-52.1545 ■	-79.6209 ■	-16.4626 ★	-247.2897 ■	-303.5774 ■
19	-11.2733 ★	-73.0209 ■	-123.6302 ■	-25.9890 ★	-281.2977 ■	<-307.6527 ■
18	-12.7018 ★	-88.6692 ■	-147.0561 ■	-28.3463 ★	-302.9424 ■	<-307.6527 ■
17	-12.6835 ★	-111.6754 ■	-166.8307 ■	-33.8601 ★	<-307.6527 ■	<-307.6527 ■
16	-17.7109 ★	-129.2479 ■	-198.3828 ■	-39.1269 ★	<-307.6527 ■	<-307.6527 ■
15	-15.6675 ★	-136.5161 ■	-197.7399 ■	-36.7544 ★	<-307.6527 ■	<-307.6527 ■
14	-10.5037 ★	-141.9135 ■	-190.2602 ■	-29.8205 ★	<-307.6527 ■	<-307.6527 ■
13	-10.1705 ★	-144.1915 ■	-193.5019 ■	-33.5294 ★	<-307.6527 ■	<-307.6527 ■
12	-8.8186 ★	-128.8481 ■	-173.6689 ■	-35.6342 ★	<-307.6527 ■	<-307.6527 ■
11	-5.5186 ★	-107.6806 ■	-134.8406 ■	-26.1707 ★	<-307.6527 ■	<-307.6527 ■
10	-4.0233 ★	-83.6869 ■	-104.4692 ■	-23.5590 ★	<-307.6527 ■	<-307.6527 ■
9	-1.7330 –	-45.3637 ■	-53.6469 ■	-18.6183 ★	-272.6245 ■	<-307.6527 ■
8	-0.8745 –	-21.5054 ■	-15.9153 ■	-14.2856 ★	-213.3631 ■	-254.8639 ■
7	-8.1100 ♦	-4.4342 ■	-0.0398 –	-8.7902 ★	-144.4376 ■	-183.3693 ■
6	-11.7983 ♦	-0.3895 –	-8.0615 ♦	-14.3916 ★	-91.2397 ■	-144.8587 ■
5	-14.0106 ♦	-9.4543 ★	-29.0700 ♦	-21.0804 ★	-79.8080 ■	-146.3293 ■
4	-17.8780 ♦	-35.8601 ★	-68.2808 ♦	-23.2754 ★	-96.8229 ■	-163.9219 ■
3	-15.5479 ♦	-91.9389 ★	-123.9708 ♦	-22.4523 ★	-132.7415 ■	-199.4006 ■
2	-8.1596 ♦	-169.5987 ★	-194.1892 ♦	-5.3108 ★	-105.5927 ■	-129.3457 ■

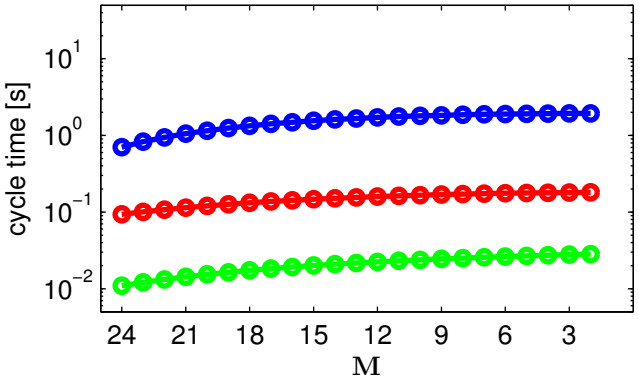


Figure 14. Gamma Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

7) Inverse Gamma Distribution: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \,\mathrm{IGamma}(x; \alpha_j^I, \beta_j^I), \text{ where,} \tag{43a}$$

$$\mathrm{IGamma}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right), \tag{43b}$$

$$N = 25, \tag{43c}$$

$$\alpha_j^I \sim \mathcal{U}(5, 25), \tag{43d}$$

$$\beta_j^I \sim \mathcal{U}(1, 10), \tag{43e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{43f}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 15.

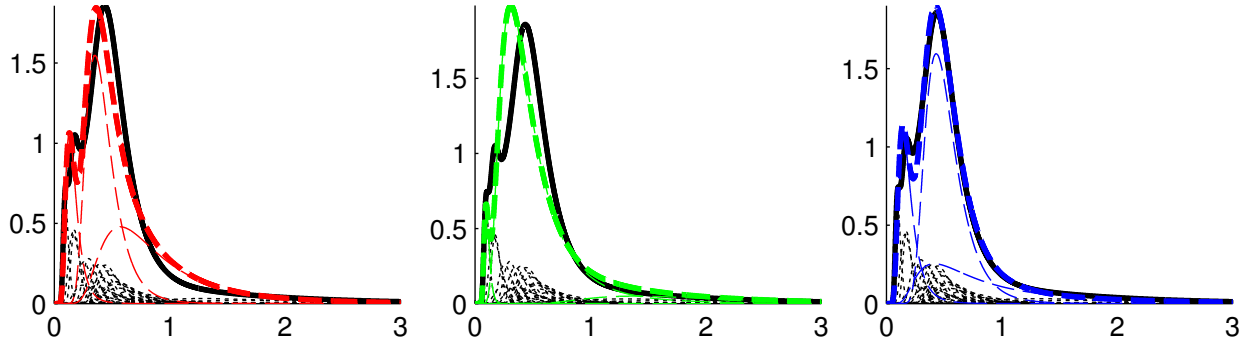


Figure 15. Mixture of Inverse Gamma Distribution: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm.

The pairwise comparison of the MRAs in the MC simulation is given in Table IX. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Similar to the gamma mixture densities, the best MRA for an application should be selected based on the reduction aggressiveness used in the application. ISE outperforms AKL and SKL even in terms of KLD between the original mixture and its approximate density. SKL is just a heuristic, while ISE approach for inverse gamma mixture densities can be evaluated analytically and the only approximation in this approach is due to the greedy implementation of it. In contrast, the AKL approach involves two approximations; the first approximation is due to the fact that AKL uses an upper bound on the KLD instead of its exact values and the second approximation is due to the greedy implementation of the reduction algorithm.

Table IX

INVERSE GAMMA DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.1609 –	-0.5311 –	-0.8694 –	-1.2329 –	-26.2719 ■	-33.3529 ■
23	-0.7481 –	-2.8188 ■	-5.0582 ■	-2.4316 ★	-71.3995 ■	-86.0777 ■
22	-0.4179 –	-3.3223 ■	-4.1833 ■	-5.5649 ★	-94.8292 ■	-120.1208 ■
21	-1.5796 –	-6.1442 ■	-11.8481 ■	-5.0535 ★	-135.4587 ■	-157.1068 ■
20	-3.7353 ★	-9.7132 ■	-22.6727 ■	-7.7211 ★	-157.1496 ■	-186.4425 ■
19	-3.1680 ★	-20.8785 ■	-34.8160 ■	-8.7032 ★	-183.9012 ■	-209.0102 ■
18	-5.0317 ★	-33.7120 ■	-55.9494 ■	-11.1149 ★	-188.4699 ■	-228.4490 ■
17	-8.0720 ★	-49.8177 ■	-83.1881 ■	-10.3933 ★	-206.9608 ■	-235.3128 ■
16	-8.1754 ★	-62.2370 ■	-95.7452 ■	-10.6356 ★	-217.1249 ■	-246.6769 ■
15	-9.4291 ★	-74.9577 ■	-112.7746 ■	-10.2575 ★	-225.4618 ■	-252.3691 ■
14	-10.4917 ★	-86.4313 ■	-122.5594 ■	-13.0344 ★	-224.3687 ■	-259.2674 ■
13	-10.0456 ★	-95.7361 ■	-129.8104 ■	-11.8534 ★	-227.2092 ■	-251.4698 ■
12	-9.6704 ★	-99.1614 ■	-135.8530 ■	-11.2544 ★	-230.9819 ■	-254.5023 ■
11	-5.2657 ★	-108.8319 ■	-134.6433 ■	-7.9139 ★	-225.2311 ■	-248.8233 ■
10	-5.3124 ★	-96.7942 ■	-120.2859 ■	-7.9953 ★	-215.7203 ■	-235.4470 ■
9	-4.3765 ★	-86.7625 ■	-106.2182 ■	-6.6277 ★	-191.0635 ■	-212.9594 ■
8	-2.0210 ★	-70.4941 ■	-81.5947 ■	-4.5170 ★	-164.1391 ■	-182.8441 ■
7	-0.4355 –	-45.9892 ■	-50.4595 ■	-2.3818 ★	-137.8935 ■	-153.1023 ■
6	-1.1359 –	-30.3020 ■	-23.4249 ■	-0.5178 –	-95.4391 ■	-102.3420 ■
5	-4.2514 ♦	-19.4156 ■	-8.0275 ■	-0.5169 –	-66.4868 ■	-58.5799 ■
4	-7.4272 ♦	-14.7045 ■	-2.8237 ■	-1.6759 –	-47.6291 ■	-36.3498 ■
3	-12.3307 ♦	-8.1202 ■	-0.2079 –	-9.8105 ♦	-56.5208 ■	-21.9279 ■
2	-10.0884 ♦	-1.2262 –	-2.8988 ♦	-13.8318 ♦	-43.7782 ■	-9.9918 ■

The average cycle times for MRAs in the MC simulations are given in Figure 16. The ISE is the most costly MRA and SKL is the least costly MRA.

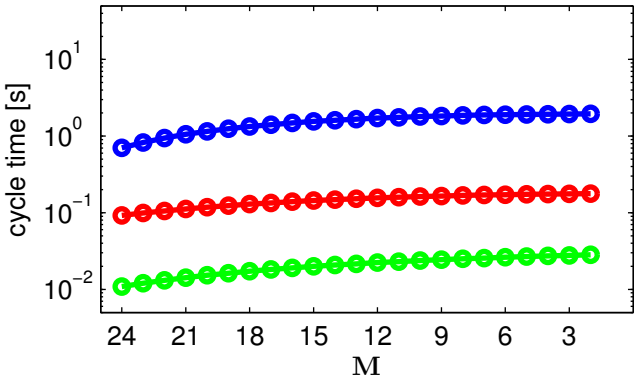


Figure 16. Inverse Gamma Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

8) Univariate Gaussian Distribution: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \mathcal{N}(x; \mu_j^I, (\sigma_j^I)^2), \text{ where,} \tag{44a}$$

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \tag{44b}$$

$$N = 25, \tag{44c}$$

$$\mu_j^I \sim \mathcal{U}(0, 100), \tag{44d}$$

$$(\sigma_j^I)^2 \sim \mathcal{U}(0.05, 4.05), \tag{44e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{44f}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x)$.

One of the mixture density functions generated in the MC simulation along with its reduced approximations with 3 components using three reduction algorithms AKL, SKL and ISE are plotted in Figure 17.

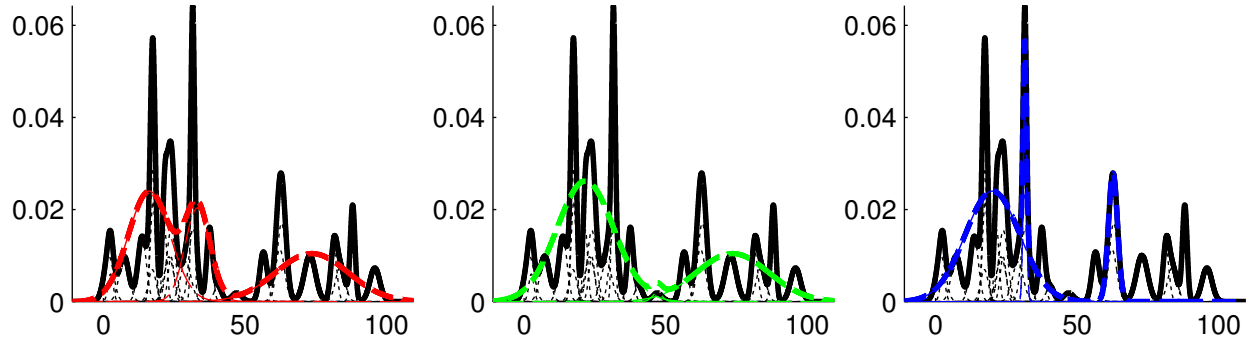


Figure 17. Mixture of Univariate Gaussian Distribution: A realization of the original mixture density and its approximations are illustrated. The original mixture density (black solid line) and its components (black dashed line) are given. In the sub-figures AKL, SKL and ISE are used to approximate the original mixture which has 25 component densities with mixtures with 3 component densities. The approximate densities (thick dashed lines) and their components (thin dashed line) are drawn in different colors; red(AKL), green(SKL) and blue(ISE). AKL is used in the left sub-figure, SKL is used in the center sub-figure and ISE is used in the right sub-figure. The reduced mixture in the right sub-figure is not rescaled after possible pruning steps and is plotted as it is used in the ISE algorithm.

The pairwise comparison of the MRAs in the MC simulation is given in Table X. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL especially at increased reduction aggressiveness which is where the reduction algorithms are mostly used. AKL obtains lower ISE compared to SKL which confirms that AKL is better MRA compared to SKL for Gaussian mixtures. When SKL and ISE are compared with respect to KLD between the original mixture and its approximation, as the reduction aggressiveness increases, the MRA with lower error changes to SKL from ISE. This phenomenon can be attributed to the property ISE has and does not exist in SKL. In the

ISE the mixture can be reduced via pruning, whereas in the other two algorithms pruning does not exist. Since the pruning can significantly remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, the SKL can obtain smaller KLD between the original mixture and its approximation.

Table X
UNIVARIATE GAUSSIAN DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			comparison with respect to ISE		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.4332 –	-5.2890 ■	-7.3009 ■	-0.9522 –	-9.6273 ■	-13.9983 ■
23	-0.8665 –	-17.9044 ■	-23.4943 ■	-2.5838 ★	-23.5852 ■	-36.0395 ■
22	-1.0764 –	-26.1343 ■	-33.5425 ■	-3.7856 ★	-33.8329 ■	-51.8804 ■
21	-1.4102 –	-30.7829 ■	-40.7042 ■	-5.3755 ★	-42.6739 ■	-68.1350 ■
20	-1.4697 –	-29.7295 ■	-38.4395 ■	-6.9623 ★	-53.7073 ■	-83.0968 ■
19	-1.1190 –	-23.4374 ■	-28.5975 ■	-7.1881 ★	-57.4235 ■	-86.8187 ■
18	-0.0766 –	-12.3330 ■	-11.9444 ■	-4.8056 ★	-61.2641 ■	-82.1202 ■
17	-0.0859 –	-3.1319 ■	-3.2897 ■	-6.4831 ★	-64.4430 ■	-95.3588 ■
16	-0.1969 –	-1.0359 –	-1.1656 –	-6.3125 ★	-70.7273 ■	-100.8065 ■
15	-0.5537 –	-9.4699 ★	-10.3523 ♦	-7.0545 ★	-74.6596 ■	-108.7929 ■
14	-0.8800 –	-23.8219 ★	-26.7078 ♦	-5.7127 ★	-79.2093 ■	-110.4159 ■
13	-1.7890 –	-45.5494 ★	-51.4026 ♦	-5.1574 ★	-86.1219 ■	-115.3644 ■
12	-0.5516 –	-77.6616 ★	-79.9074 ♦	-8.2813 ★	-91.7188 ■	-124.1990 ■
11	-0.1579 –	-106.3933 ★	-102.5797 ♦	-12.0241 ★	-95.6732 ■	-137.6829 ■
10	-1.3398 –	-124.2827 ★	-114.1771 ♦	-17.4090 ★	-101.1150 ■	-154.7190 ■
9	-5.9391 ★	-153.7178 ★	-130.4077 ♦	-28.0047 ★	-102.8061 ■	-180.1624 ■
8	-17.7850 ★	-176.8338 ★	-140.6988 ♦	-39.7180 ★	-104.5158 ■	-193.1350 ■
7	-39.5323 ★	-200.2816 ★	-143.1803 ♦	-60.3924 ★	-106.3392 ■	-220.6487 ■
6	-70.2568 ★	-210.7818 ★	-132.8848 ♦	-77.3235 ★	-103.1100 ■	-232.5035 ■
5	-112.2149 ★	-200.8386 ★	-96.7024 ♦	-97.9758 ★	-89.5019 ■	-245.8193 ■
4	-148.4252 ★	-178.4029 ★	-59.1499 ♦	-104.0632 ★	-71.0769 ■	-240.2691 ■
3	-151.9499 ★	-150.8297 ★	-38.0336 ♦	-69.7368 ★	-54.7892 ■	-193.8287 ■
2	-98.9634 ★	-109.4630 ★	-27.7333 ♦	-25.3379 ★	-19.1444 ■	-83.3154 ■

The average cycle times for MRAs in the MC simulations are given in Figure 18. The ISE is the most costly MRA and SKL is the least costly MRA.

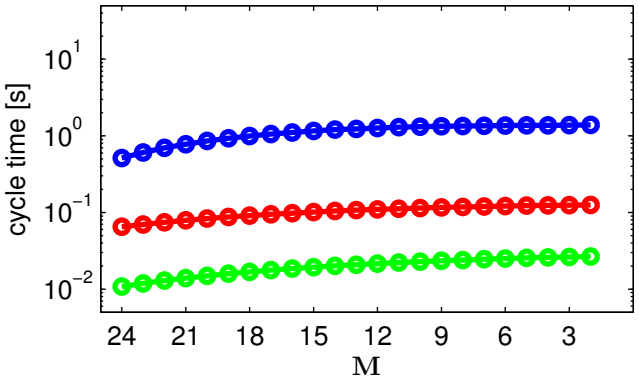


Figure 18. Univariate Gaussian Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

9) Multivariate Gaussian Distribution: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \mathcal{N}(x; m_j^I, P_j^I), \text{ where,} \tag{45a}$$

$$\mathcal{N}(x; m, P) = (2\pi)^{-\frac{k}{2}} |P|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-m)^{\mathrm{T}} P^{-1}(x-m)\right), \tag{45b}$$

$$N = 25, \tag{45c}$$

$$m_j^I = [m_{j,1}^I, m_{j,2}^I]^{\mathrm{T}} \text{ and } m_{j,1}^I, m_{j,2}^I \sim \mathcal{U}(0, 40), \tag{45d}$$

$$P_j^I = 0.1 I_2 + \left(E_j^I + (E_j^I)^{\mathrm{T}}\right)^2, \text{ where elements of the matrix } E, E_{uv} \sim \mathcal{U}(0,1), \tag{45e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1), \tag{45f}$$

where, I_2 is the 2×2 identity matrix.

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^4$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{x_j^r\}_{r=1}^S \overset{\text{iid}}{\sim} p_j(x)$. The reduced number of samples is due to the longer simulation time for the multivariate distributions.

The pairwise comparison of the MRAs in the MC simulation is given in Table XI. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL. AKL obtains lower ISE compared to SKL which confirms that AKL is better MRA compared to SKL for Gaussian mixtures. When SKL and ISE are compared with respect to KLD between the original mixture and its approximation, the SKL obtains lower KLD values. This observation can be due to the fact that in the ISE the mixture can be reduced via pruning, whereas in the other two algorithms pruning does not exist. Since the pruning can significantly remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, the SKL can obtain smaller KLD between the original mixture and its approximation.

Table XI
MULTIVARIATE GAUSSIAN DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			comparison with respect to ISE		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-1.1945 –	-35.9579 ★	-28.9447 ♦	-14.2602 ★	-24.9937 ■	-63.4507 ■
23	-1.1600 –	-144.9496 ★	-135.8489 ♦	-36.4474 ★	-75.6239 ■	-150.6456 ■
22	-0.5908 –	-221.5719 ★	-215.7150 ♦	-58.4619 ★	-109.7033 ■	-207.4378 ■
21	-0.6233 –	-280.1105 ★	-276.8938 ♦	-81.4599 ★	-149.0404 ■	-251.8204 ■
20	-0.6489 –	-305.4287 ★	-303.8165 ♦	-109.4950 ★	-169.3420 ■	-276.5285 ■
19	-0.5871 –	<-307.6527 ★	<-307.6527 ♦	-121.8904 ★	-189.6884 ■	-293.5446 ■
18	-0.7244 –	<-307.6527 ★	<-307.6527 ♦	-128.4476 ★	-209.6783 ■	-300.6798 ■
17	-1.4344 –	<-307.6527 ★	<-307.6527 ♦	-139.2251 ★	-226.1157 ■	-307.4682 ■
16	-1.7705 –	<-307.6527 ★	<-307.6527 ♦	-140.7026 ★	-241.1962 ■	<-307.6527 ■
15	-3.4139 ★	<-307.6527 ★	<-307.6527 ♦	-141.7236 ★	-255.2380 ■	<-307.6527 ■
14	-4.5821 ★	<-307.6527 ★	<-307.6527 ♦	-138.9169 ★	-263.6487 ■	<-307.6527 ■
13	-7.4590 ★	<-307.6527 ★	<-307.6527 ♦	-140.3553 ★	-272.5884 ■	<-307.6527 ■
12	-12.0023 ★	<-307.6527 ★	<-307.6527 ♦	-138.5765 ★	-276.4985 ■	<-307.6527 ■
11	-19.5781 ★	<-307.6527 ★	<-307.6527 ♦	-125.4307 ★	-280.9759 ■	<-307.6527 ■
10	-32.6070 ★	<-307.6527 ★	<-307.6527 ♦	-114.7964 ★	-282.7278 ■	<-307.6527 ■
9	-46.8846 ★	<-307.6527 ★	<-307.6527 ♦	-97.2822 ★	-287.7023 ■	<-307.6527 ■
8	-66.1622 ★	<-307.6527 ★	<-307.6527 ♦	-78.0591 ★	-288.7391 ■	<-307.6527 ■
7	-87.0650 ★	<-307.6527 ★	<-307.6527 ♦	-59.2948 ★	-287.4202 ■	<-307.6527 ■
6	-104.2347 ★	<-307.6527 ★	<-307.6527 ♦	-42.1362 ★	-289.0317 ■	<-307.6527 ■
5	-108.9983 ★	<-307.6527 ★	<-307.6527 ♦	-26.4175 ★	-279.2933 ■	<-307.6527 ■
4	-102.8250 ★	<-307.6527 ★	<-307.6527 ♦	-14.2477 ★	-260.5524 ■	-294.0329 ■
3	-71.3789 ★	<-307.6527 ★	<-307.6527 ♦	-4.9764 ★	-218.0598 ■	-250.0571 ■
2	-25.0232 ★	<-307.6527 ★	<-307.6527 ♦	-1.5960 –	-136.6015 ■	-156.8068 ■

The average cycle times for MRAs in the MC simulations are given in Figure 19. The ISE is the most costly MRA and SKL is the least costly MRA.

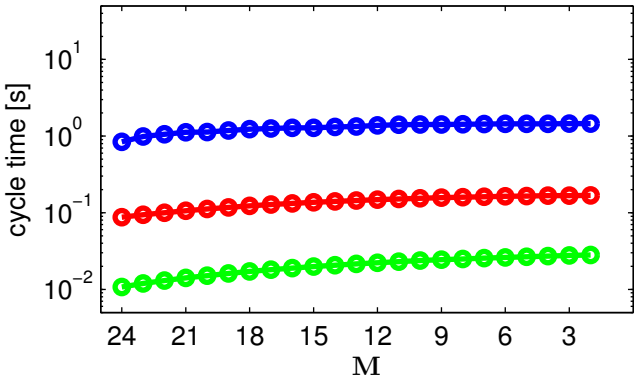


Figure 19. Multivariate Gaussian Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

10) Gaussian Gamma Distribution: For the j^{th} MC run the mixture density function $p_j(x,\tau)$ is selected as

$$p_j(x,\tau) = \sum_{I=1}^{N} w_j^I \,\text{GaussianGamma}\left(x,\tau;\mu_j^I,\lambda_j^I,\alpha_j^I,\beta_j^I\right), \text{ where,} \tag{46a}$$

$$\text{GaussianGamma}\left(x,\tau;\mu,\lambda,\alpha,\beta\right) = \mathcal{N}\left(x;\mu,\frac{1}{\lambda\tau}\right)\text{Gamma}(\tau;\alpha,\beta), \tag{46b}$$

$$N = 25, \tag{46c}$$

$$\mu_j^I \sim \mathcal{U}(0,50), \tag{46d}$$

$$\frac{1}{\lambda_j^I} \sim \mathcal{U}(0.1,10.1), \tag{46e}$$

$$\alpha_j^I \sim \mathcal{U}(10,110), \tag{46f}$$

$$\beta_j^I \sim \mathcal{U}(1,101), \tag{46g}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1,1.1). \tag{46h}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^6$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{(x_j^r,\tau_j^r)\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x,\tau)$.

The pairwise comparison of the MRAs in the MC simulation is given in Table XII. The simulation results for Gaussian gamma distribution in Table XII is similar to the simulation results for the univariate Gaussian distribution in Table X and can be explained similarly.

The average cycle times for MRAs in the MC simulations are given in Figure 20. The ISE is the most costly MRA and SKL is the least costly MRA.

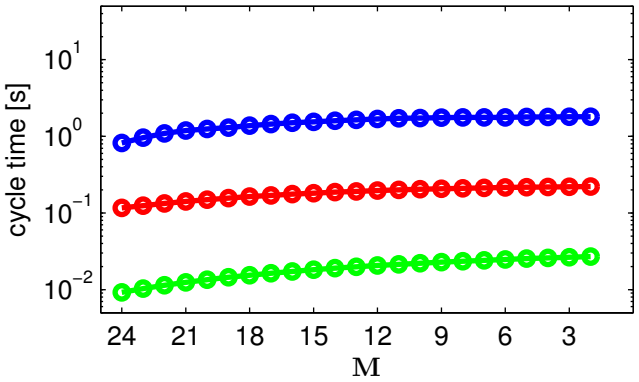


Figure 20. Gaussian Gamma Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

Table XII
GAUSSIAN GAMMA DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-227.6436 ★	-71.5902 ■	-266.5589 ■	-159.4115 ★	-122.9047 ■	<-307.6527 ■
23	-287.3936 ★	-51.0382 ■	-240.1488 ■	-217.6828 ★	-201.1852 ■	<-307.6527 ■
22	<-307.6527 ★	-12.8720 ■	-207.8191 ■	-252.9553 ★	-247.4863 ■	<-307.6527 ■
21	<-307.6527 ★	-0.0799 –	-182.7460 ■	-270.3675 ★	-284.5875 ■	<-307.6527 ■
20	<-307.6527 ★	-9.0149 ★	-140.9992 ■	-285.3047 ★	-304.7125 ■	<-307.6527 ■
19	<-307.6527 ★	-33.6319 ★	-100.9815 ■	-294.3741 ★	-305.5870 ■	<-307.6527 ■
18	<-307.6527 ★	-80.0531 ★	-49.2776 ■	-299.3078 ★	<-307.6527 ■	<-307.6527 ■
17	<-307.6527 ★	-119.4771 ★	-17.4956 ■	-300.2106 ★	<-307.6527 ■	<-307.6527 ■
16	<-307.6527 ★	-151.1616 ★	-1.6034 –	-297.6939 ★	<-307.6527 ■	<-307.6527 ■
15	<-307.6527 ★	-185.5364 ★	-9.3191 ♦	-292.1848 ★	<-307.6527 ■	<-307.6527 ■
14	<-307.6527 ★	-207.5194 ★	-32.7559 ♦	-287.4913 ★	<-307.6527 ■	<-307.6527 ■
13	-304.1922 ★	-229.0621 ★	-82.8152 ♦	-279.7282 ★	<-307.6527 ■	<-307.6527 ■
12	-290.4785 ★	-260.5154 ★	-139.9672 ♦	-269.0099 ★	<-307.6527 ■	<-307.6527 ■
11	-273.5811 ★	-273.8682 ★	-179.1024 ♦	-258.3999 ★	<-307.6527 ■	<-307.6527 ■
10	-248.6395 ★	-280.5020 ★	-217.2837 ♦	-244.8790 ★	<-307.6527 ■	<-307.6527 ■
9	-216.5246 ★	-294.9255 ★	-245.1078 ♦	-223.4275 ★	<-307.6527 ■	<-307.6527 ■
8	-181.2870 ★	-304.9163 ★	-271.2231 ♦	-198.6707 ★	<-307.6527 ■	<-307.6527 ■
7	-137.9164 ★	<-307.6527 ★	-291.6520 ♦	-159.6304 ★	<-307.6527 ■	<-307.6527 ■
6	-98.2444 ★	<-307.6527 ★	-294.5075 ♦	-117.1293 ★	<-307.6527 ■	<-307.6527 ■
5	-64.3455 ★	<-307.6527 ★	-303.2972 ♦	-75.9593 ★	<-307.6527 ■	<-307.6527 ■
4	-39.7933 ★	<-307.6527 ★	-300.9611 ♦	-42.7454 ★	-303.5481 ■	<-307.6527 ■
3	-20.0898 ★	-305.2584 ★	-295.9839 ♦	-22.0717 ★	-266.1945 ■	-294.9601 ■
2	-6.0898 ★	-294.7226 ★	-290.1313 ♦	-6.8558 ★	-180.9994 ■	-212.3812 ■

11) Dirichlet Distribution: For the j^{th} MC run the mixture density function $p_j(x)$ is selected as

$$p_j(x) = \sum_{I=1}^{N} w_j^I \operatorname{Dir}_K(x; \alpha_j^I), \text{ where,} \tag{47a}$$

$$\operatorname{Dir}_K(x; \alpha) = \frac{1}{\mathrm{B}(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}, \tag{47b}$$

$$N = 25, \tag{47c}$$

$$K = 3, \tag{47d}$$

$$\alpha_{i,j}^I \sim \mathcal{U}(3, 33), \tag{47e}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I \right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{47f}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, a two dimensional uniform grid with 1000 discretization points on each dimension is used.

The pairwise comparison of the MRAs in the MC simulation is given in Table XIII. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL at increased reduction aggressiveness where the reduction algorithms are mostly used. AKL obtains lower ISE compared to SKL, which confirms that AKL is better MRA compared to SKL for Dirichlet distribution. When SKL and ISE are compared with respect to KLD between the original mixture and its approximation at increased reduction aggressiveness, the SKL obtains lower KLD values. This can be due to the fact that in the ISE the mixture can be reduced via pruning, whereas in the other two algorithms pruning does not exist. Since the pruning can significantly remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, the SKL can obtain smaller KLD between the original mixture and its approximation.

The average cycle times for MRAs in the MC simulations are given in Figure 21. The ISE is the most costly MRA and SKL is the least costly MRA.

Table XIII
DIRICHLET DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			comparison with respect to ISE		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-2.5666 ★	-0.2398 –	-3.3628 ■	-4.3973 ★	-0.8360 –	-7.3809 ■
23	-7.4912 ★	-0.7447 –	-10.9984 ■	-13.6783 ★	-2.6157 ■	-24.7268 ■
22	-10.3647 ★	-1.0947 –	-16.0159 ■	-21.4649 ★	-4.7341 ■	-41.3870 ■
21	-9.6935 ★	-1.8419 –	-17.6096 ■	-30.3860 ★	-7.6855 ■	-60.8076 ■
20	-12.4159 ★	-2.1524 ■	-21.8658 ■	-38.1783 ★	-12.0844 ■	-78.8719 ■
19	-11.5749 ★	-2.5228 ■	-21.6589 ■	-46.1438 ★	-14.7645 ■	-94.1257 ■
18	-9.7118 ★	-2.3334 ■	-19.0432 ■	-46.4753 ★	-16.6976 ■	-98.5374 ■
17	-8.6906 ★	-4.0209 ■	-20.5934 ■	-49.8952 ★	-20.0568 ■	-108.9522 ■
16	-9.6885 ★	-3.5472 ■	-21.0923 ■	-54.8580 ★	-22.5279 ■	-117.7343 ■
15	-6.5961 ★	-3.6113 ■	-17.1755 ■	-51.5101 ★	-24.1902 ■	-117.2479 ■
14	-5.1010 ★	-2.6479 ■	-12.0721 ■	-51.7083 ★	-25.6784 ■	-117.9129 ■
13	-2.8814 ★	-1.3254 –	-6.2265 ■	-49.5491 ★	-26.7020 ■	-118.7117 ■
12	-2.1380 ★	-1.3019 –	-4.8947 ■	-48.8404 ★	-29.3112 ■	-119.4869 ■
11	-2.1684 ★	-0.7470 –	-3.8610 ■	-55.6602 ★	-29.8666 ■	-133.7016 ■
10	-0.6738 –	-0.0228 –	-0.5464 –	-48.6114 ★	-31.2997 ■	-127.2023 ■
9	-0.5030 –	-0.0781 –	-0.3589 –	-53.0079 ★	-31.8677 ■	-131.9862 ■
8	-0.0441 –	-1.1542 –	-0.9927 –	-50.1513 ★	-33.8660 ■	-126.3546 ■
7	-0.2928 –	-1.8165 –	-1.2940 –	-55.4659 ★	-36.4378 ■	-130.1813 ■
6	-1.0472 –	-4.2325 ★	-2.2235 ♦	-63.0756 ★	-37.5934 ■	-140.3286 ■
5	-2.4963 ★	-11.3319 ★	-5.8765 ♦	-62.1045 ★	-38.8539 ■	-138.8603 ■
4	-7.8057 ★	-17.2280 ★	-5.5499 ♦	-65.5619 ★	-46.4077 ■	-151.9018 ■
3	-15.3071 ★	-22.6518 ★	-4.9358 ♦	-60.3392 ★	-43.1332 ■	-147.1539 ■
2	-8.5852 ★	-21.2155 ★	-6.9963 ♦	-22.9241 ★	-25.2177 ■	-79.2735 ■

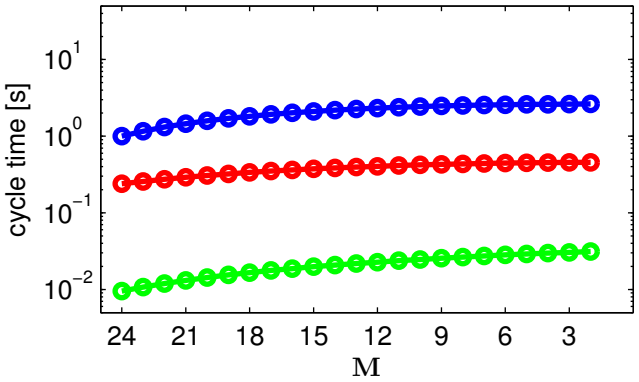


Figure 21. Dirichlet Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

12) Wishart Distribution: For the j^{th} MC run the mixture density function $p_j(X)$ is selected as

$$p_j(X) = \sum_{I=1}^{N} w_j^I \mathcal{W}_d(X; n_j^I, V_j^I), \text{ where,} \tag{48a}$$

$$\mathcal{W}_d(X; n, V) = \frac{|X|^{\frac{1}{2}(n-d-1)} \exp \mathrm{Tr}\left(-\frac{1}{2}V^{-1}X\right)}{2^{\frac{1}{2}nd}\Gamma_d(\frac{1}{2}n)|V|^{\frac{1}{2}n}}, \tag{48b}$$

$$N = 25, \tag{48c}$$

$$d = 2, \tag{48d}$$

$$n_j^I \sim \mathcal{U}(10, 20), \tag{48e}$$

$$V_j^I = I_2 + (E_j^I + (E_j^I)^{\mathrm{T}})^2, \text{ where elements of the matrix } E, E_{uv} \sim \mathcal{U}(0, 4), \tag{48f}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{48g}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^4$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{X_j^r\}_{r=1}^S \overset{\text{iid}}{\sim} p_j(X)$. The reduced number of samples is due to the longer simulation time for the matrix-variate distributions.

The pairwise comparison of the MRAs in the MC simulation is given in Table XIV. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL. AKL obtains lower ISE compared to SKL. When SKL and ISE are compared with respect to KLD between the original mixture and its approximation, the SKL obtains lower KLD values. This can be due to the fact that in the ISE, the mixture can be reduced via pruning, whereas in the other two algorithms pruning does not exist. Since the pruning can significantly remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, the SKL can obtain smaller KLD between the original mixture and its approximation.

Table XIV
WISHART DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-2.0037 ★	-15.3982 ★	-8.3128 ♦	-2.9121 ★	-18.5149 ■	-33.0444 ■
23	-2.8113 ★	-38.6435 ★	-25.6197 ♦	-5.1595 ★	-52.6338 ■	-81.3282 ■
22	-4.5437 ★	-58.7701 ★	-38.3541 ♦	-5.3312 ★	-87.2254 ■	-115.7220 ■
21	-6.8043 ★	-86.7639 ★	-60.2630 ♦	-5.5140 ★	-109.3612 ■	-138.1363 ■
20	-9.2712 ★	-99.0152 ★	-69.1757 ♦	-6.1554 ★	-131.8519 ■	-163.2915 ■
19	-8.5149 ★	-119.6872 ★	-90.7910 ♦	-7.6318 ★	-157.1708 ■	-184.0388 ■
18	-8.3739 ★	-140.3876 ★	-112.0419 ♦	-5.2267 ★	-181.0762 ■	-199.1293 ■
17	-8.6371 ★	-156.2561 ★	-125.7009 ♦	-4.5949 ★	-196.8157 ■	-213.9674 ■
16	-10.4086 ★	-172.0356 ★	-137.7513 ♦	-3.9876 ★	-200.9610 ■	-223.4978 ■
15	-10.8153 ★	-183.0596 ★	-149.3940 ♦	-4.3739 ★	-213.7053 ■	-232.0943 ■
14	-14.9880 ★	-185.4108 ★	-148.1958 ♦	-2.8692 ★	-222.9746 ■	-238.2579 ■
13	-17.5781 ★	-201.1911 ★	-159.7920 ♦	-2.0920 ★	-232.1875 ■	-241.1627 ■
12	-20.0245 ★	-210.0494 ★	-165.3356 ♦	-0.8004 –	-238.1412 ■	-238.8539 ■
11	-25.8060 ★	-215.3472 ★	-165.5753 ♦	-1.2169 –	-237.9291 ■	-247.8407 ■
10	-30.4702 ★	-222.1627 ★	-166.8207 ♦	-0.8837 –	-244.8608 ■	-243.2243 ■
9	-37.6224 ★	-214.2754 ★	-147.4754 ♦	-0.7677 –	-242.9515 ■	-244.1493 ■
8	-45.9761 ★	-207.2280 ★	-132.9992 ♦	-0.4768 –	-238.8728 ■	-242.7070 ■
7	-56.0072 ★	-200.0820 ★	-115.6448 ♦	-0.1090 –	-232.0307 ■	-237.9346 ■
6	-70.3193 ★	-196.9611 ★	-106.3166 ♦	-0.1781 –	-224.9114 ■	-226.5835 ■
5	-87.7632 ★	-199.0765 ★	-95.5556 ♦	-0.0180 –	-213.1065 ■	-217.9067 ■
4	-96.5746 ★	-191.3252 ★	-88.5331 ♦	-0.1093 –	-190.6748 ■	-202.0341 ■
3	-99.0295 ★	-187.0054 ★	-99.5397 ♦	-0.4120 –	-153.4433 ■	-173.0074 ■
2	-60.6736 ★	-183.9549 ★	-130.0349 ♦	-0.8849 –	-96.6770 ■	-119.4425 ■

The average cycle times for MRAs in the MC simulations are given in Figure 22. The ISE is the most costly MRA and SKL is the least costly MRA.

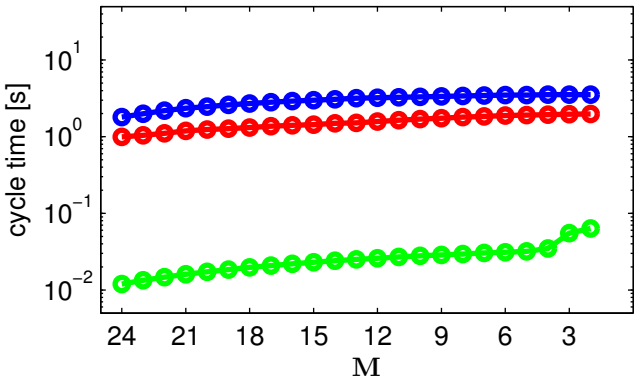


Figure 22. Wishart Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

13) Inverse Wishart Distribution: For the j^{th} MC run the mixture density function $p_j(X)$ is selected as

$$p_j(X) = \sum_{I=1}^{N} w_j^I \mathcal{IW}_d(X; \nu_j^I, \Psi_j^I), \text{ where,} \tag{49a}$$

$$\mathcal{IW}_d(X; \nu, \Psi) = \frac{|\Psi|^{\frac{1}{2}(\nu-d-1)} \exp \operatorname{Tr}\left(-\frac{1}{2}\Psi X^{-1}\right)}{2^{\frac{1}{2}(\nu-d-1)d}\Gamma_d\left[\frac{1}{2}(\nu-d-1)\right]|X|^{\frac{\nu}{2}}}, \tag{49b}$$

$$N = 25, \tag{49c}$$

$$d = 2, \tag{49d}$$

$$\nu_j^I \sim \mathcal{U}(20, 120), \tag{49e}$$

$$\Psi_j^I = 0.1 I_2 + (E_j^I + (E_j^I)^{\mathrm{T}})^2, \text{ where elements of the matrix } E, E_{uv} \sim \mathcal{U}(0, 20), \tag{49f}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{49g}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^4$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{X_j^r\}_{r=1}^S \overset{\text{iid}}{\sim} p_j(X)$. The reduced number of samples is due to the longer simulation time for the matrix-variate distributions.

The pairwise comparison of the MRAs in the MC simulation is given in Table XV. The simulation results for inverse Wishart distribution in Table XV are different from the simulation results for the Wishart distribution in Table XIV in columns two to four. A closer look at the simulation results reveals that ISE reduces the inverse Wishart mixture densities mostly via pruning which is due to the larger spread of the mixture parameters compared to the Wishart mixture density. Since the pruning can significantly remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, AKL and SKL can obtain smaller KLD between the original mixture and its approximation especially when the reduction aggressiveness is increased.

Table XV
INVERSE WISHART DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			**comparison with respect to ISE**		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-0.3822 –	-31.0681 ■	-30.6881 ■	-2.7842 ★	-123.8729 ■	-143.0014 ■
23	-0.4998 –	-35.9238 ■	-37.6876 ■	-1.6896 –	-208.4300 ■	-214.7503 ■
22	-0.6540 –	-36.5143 ■	-40.8013 ■	-1.2811 –	-254.3702 ■	-252.5808 ■
21	-0.9582 –	-34.3612 ■	-41.1308 ■	-0.3515 –	-274.2704 ■	-273.4451 ■
20	-2.2109 ★	-30.2270 ■	-40.8530 ■	-0.4774 –	-286.3297 ■	-277.2858 ■
19	-2.5419 ★	-26.3217 ■	-40.9835 ■	-4.3904 ♦	-296.2179 ■	-288.1484 ■
18	-1.9515 –	-24.1616 ■	-38.9457 ■	-10.9292 ♦	-299.3213 ■	-287.0736 ■
17	-1.6045 –	-22.1694 ■	-37.2299 ■	-15.3868 ♦	-304.6369 ■	-287.6022 ■
16	-2.1376 ★	-18.3786 ■	-34.0799 ■	-20.4723 ♦	-305.5842 ■	-283.9304 ■
15	-2.5235 ★	-16.3839 ■	-33.1619 ■	-24.0786 ♦	-304.9361 ■	-283.3373 ■
14	-5.1973 ★	-10.8122 ■	-32.2421 ■	-28.2861 ♦	-306.6794 ■	-280.9361 ■
13	-6.1012 ★	-7.4038 ■	-27.9481 ■	-32.5216 ♦	-306.8509 ■	-277.0808 ■
12	-8.8367 ★	-4.6772 ■	-27.9276 ■	-37.3854 ♦	-303.8755 ■	-272.0584 ■
11	-11.0977 ★	-1.2811 –	-20.6372 ■	-40.9725 ♦	-303.2706 ■	-271.0462 ■
10	-9.8336 ★	-0.1287 –	-12.9356 ■	-47.2153 ♦	-304.6293 ■	-264.8871 ■
9	-10.8323 ★	-0.3292 –	-10.9327 ■	-46.0062 ♦	-300.1093 ■	-256.7529 ■
8	-12.4369 ★	-1.3110 –	-9.1348 ■	-46.7246 ♦	-298.6612 ■	-252.3657 ■
7	-14.9143 ★	-4.3483 ★	-5.7644 ■	-46.0989 ♦	-293.6502 ■	-243.5826 ■
6	-20.5298 ★	-11.3825 ★	-3.4175 ■	-47.1829 ♦	-289.1333 ■	-240.3294 ■
5	-23.9783 ★	-24.8776 ★	-0.2544 –	-47.8436 ♦	-282.9144 ■	-227.9824 ■
4	-28.1708 ★	-48.6175 ★	-3.4935 ♦	-50.0911 ♦	-281.3901 ■	-225.4796 ■
3	-41.2037 ★	-104.3283 ★	-20.3655 ♦	-53.2623 ♦	-276.4542 ■	-224.8164 ■
2	-43.2028 ★	-170.3032 ★	-67.7681 ♦	-53.5171 ♦	-271.1718 ■	-220.2100 ■

The average cycle times for MRAs in the MC simulations are given in Figure 23. SKL is the least costly MRA but ISE has almost the same cost as the AKL. A closer look at the simulation results shows that ISE reduces the inverse Wishart mixture densities in this simulation mostly via pruning. Hence, ISE does not need to calculated many new merging hypotheses which results in decreased computation time.

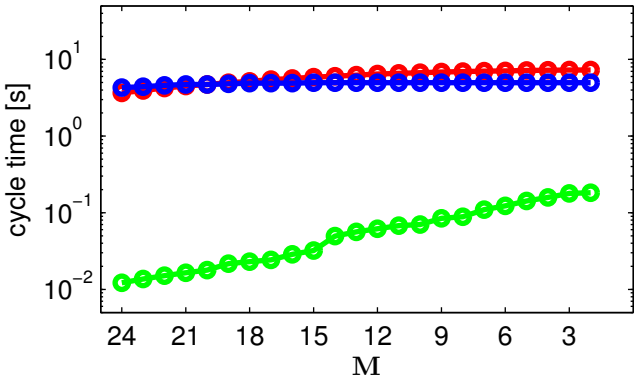


Figure 23. Inverse Wishart Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

14) Gaussian Inverse Wishart Distribution: For the j^{th} MC run the mixture density function $p_j(x, X)$ is selected as

$$p_j(x,X) = \sum_{I=1}^{N} w_j^I \,\mathrm{GIW}\left(x, X; m_j^I, P_j^I, \nu_j^I, \Psi_j^I\right), \text{ where,} \tag{50a}$$

$$\mathrm{GIW}\left(x, X; m, P, \nu, \Psi\right) = \mathcal{N}\left(x; m, P\right) \mathcal{IW}_d\left(X; \nu, \Psi\right), \tag{50b}$$

$$N = 25, \tag{50c}$$

$$d = 2, \tag{50d}$$

$$m_j^I = [m_{j,1}^I, m_{j,2}^I]^{\mathrm{T}} \text{ and } m_{j,1}^I, m_{j,2}^I \sim \mathcal{U}(0, 20), \tag{50e}$$

$$P_j^I = I_2 + (E_j^I + (E_j^I)^{\mathrm{T}})^2, \text{ where elements of the matrix } E, E_{uv} \sim \mathcal{U}(0, 2), \tag{50f}$$

$$\nu_j^I = [m_{j,1}^I, m_{j,2}^I]^{\mathrm{T}} \text{ and } m_{j,1}^I, m_{j,2}^I \sim \mathcal{U}(10, 15), \tag{50g}$$

$$\Psi_j^I = I_2 + (E_j^I + (E_j^I)^{\mathrm{T}})^2, \text{ where elements of the matrix } E, E_{uv} \sim \mathcal{U}(0, 1), \tag{50h}$$

$$w_j^I = \left(\sum_{I=1}^{N} \widehat{w}_j^I\right)^{-1} \widehat{w}_j^I, \text{ where } \widehat{w}_j^I \sim \mathcal{U}(0.1, 1.1). \tag{50i}$$

For numerical calculation of the KLD between the original mixture and the approximate mixture, $S = 10^4$ independent and identically distributed (iid) random samples are generated from the original mixture, where $\{(x_j^r, X_j^r)\}_{r=1}^{S} \overset{\text{iid}}{\sim} p_j(x, X)$. The reduced number of samples is due to the longer simulation time for the matrix-variate distributions.

The pairwise comparison of the MRAs in the MC simulation is given in Table XVI. Not surprisingly, ISE performs the best when comparison is done with respect to the ISE. Although AKL uses an upper bound instead of the exact KLD of the original mixture with its approximation, it obtains lower KLD compared to ISE and SKL especially in increased reduction aggressiveness which is where the reduction algorithms are mostly used. AKL obtains lower ISE compared to SKL. When SKL and ISE are compared with respect to KLD between the original mixture and its approximation the SKL obtains lower KLD values. This can be due to the fact that in the ISE the mixture can be reduced via pruning, whereas in the other two algorithms pruning does not exist. Since the pruning can significantly remove the probability mass from some regions in the support that had a considerable probability mass, the KLD can tend to infinity. Hence, the SKL can obtain smaller KLD between the original mixture and its approximation.

Table XVI
GAUSSIAN INVERSE WISHART DISTRIBUTION: PAIRWISE COMPARISON OF THREE REDUCTION ALGORITHMS WITH RESPECT TO THE KLD AND ISE BETWEEN THE ORIGINAL MIXTURE AND THE APPROXIMATE MIXTURE REDUCED INCREMENTALLY. THE NUMBER OF REMAINING COMPONENTS IN THE APPROXIMATE MIXTURE IS SHOWN BY **M** IN THE LEFT COLUMN. THE QUANTITY IN EACH ELEMENT IS THE DECIMAL LOGARITHM OF THE $p-$VALUE OF THE TWO SIDED WILCOXON RANK SUM TEST. WHEN THE DIFFERENCE BETWEEN THE TWO REDUCTION ALGORITHMS IS STATISTICALLY SIGNIFICANT (p-VALUE $< 1\%$) THE SYMBOL CORRESPONDING TO THE ALGORITHM WITH SMALLER MEDIAN ERROR IS GIVEN NEXT TO THE p-VALUE.

	comparison with respect to KLD			comparison with respect to ISE		
M	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**	**AKL(★)-SKL(♦)**	**AKL(★)-ISE(■)**	**SKL(♦)-ISE(■)**
24	-19.2098 ♦	-53.8247 ★	-103.8669 ♦	-77.6041 ★	-134.3236 ■	-242.5892 ■
23	-22.0844 ♦	-70.4962 ★	-115.0313 ♦	-120.3750 ★	-207.8091 ■	-304.2737 ■
22	-23.9550 ♦	-79.0603 ★	-127.0126 ♦	-135.2584 ★	-238.3752 ■	<-307.6527 ■
21	-22.4965 ♦	-83.0589 ★	-128.3095 ♦	-139.4971 ★	-258.0132 ■	<-307.6527 ■
20	-16.7435 ♦	-74.7879 ★	-113.5978 ♦	-148.2186 ★	-271.0722 ■	<-307.6527 ■
19	-15.1884 ♦	-70.5899 ★	-109.1416 ♦	-146.1049 ★	-274.8449 ■	<-307.6527 ■
18	-9.4014 ♦	-75.1842 ★	-104.5701 ♦	-147.6972 ★	-285.4919 ■	<-307.6527 ■
17	-9.0206 ♦	-76.1876 ★	-106.9969 ♦	-154.2050 ★	-291.9285 ■	<-307.6527 ■
16	-3.7001 ♦	-79.7043 ★	-97.7630 ♦	-156.3199 ★	-300.9759 ■	<-307.6527 ■
15	-0.9538 –	-89.3732 ★	-95.9047 ♦	-154.3363 ★	-305.4155 ■	<-307.6527 ■
14	-0.1980 –	-96.4829 ★	-92.0994 ♦	-155.3288 ★	-307.2308 ■	<-307.6527 ■
13	-2.0140 ★	-111.3471 ★	-91.6728 ♦	-149.9659 ★	<-307.6527 ■	<-307.6527 ■
12	-7.0247 ★	-126.0182 ★	-88.3745 ♦	-158.4520 ★	<-307.6527 ■	<-307.6527 ■
11	-11.7620 ★	-136.9387 ★	-88.3873 ♦	-158.6775 ★	<-307.6527 ■	<-307.6527 ■
10	-18.7941 ★	-150.7215 ★	-93.1378 ♦	-159.7630 ★	<-307.6527 ■	<-307.6527 ■
9	-31.5699 ★	-158.1742 ★	-83.3266 ♦	-164.9419 ★	<-307.6527 ■	<-307.6527 ■
8	-42.1038 ★	-172.1833 ★	-86.7545 ♦	-158.2348 ★	<-307.6527 ■	<-307.6527 ■
7	-52.0329 ★	-173.2488 ★	-81.4411 ♦	-163.5848 ★	<-307.6527 ■	<-307.6527 ■
6	-60.3253 ★	-179.5657 ★	-83.3456 ♦	-148.0383 ★	<-307.6527 ■	<-307.6527 ■
5	-56.4782 ★	-177.1377 ★	-84.1744 ♦	-129.0205 ★	-307.2573 ■	<-307.6527 ■
4	-47.8600 ★	-177.3733 ★	-93.1004 ♦	-107.3896 ★	-303.2205 ■	<-307.6527 ■
3	-27.9777 ★	-159.3714 ★	-99.5311 ♦	-71.4467 ★	-293.8217 ■	<-307.6527 ■
2	-7.4688 ★	-145.6603 ★	-115.9718 ♦	-30.0839 ★	-279.0562 ■	-306.2582 ■

The average cycle times for MRAs in the MC simulations are given in Figure 24. ISE and KLD are the most costly MRA and SKL is the least costly MRA. The gap between AKL and ISE is reduced compared to other mixtures which is due to the fact that ISE chooses pruning more frequently compared to univariate and multivariate distributions. After each pruning step very little computation is required to update the cost associated to each hypothesis (see section II-E).

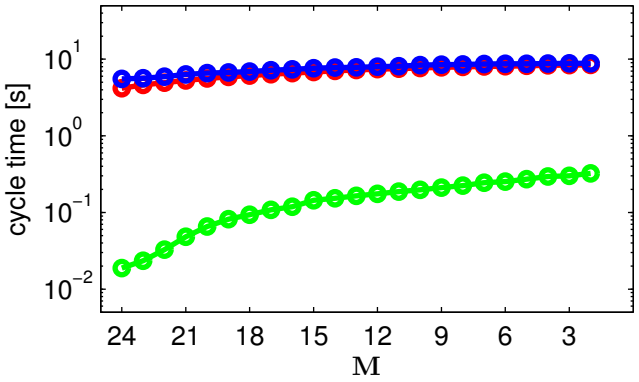


Figure 24. Gaussian Inverse Wishart Distribution: The average cycle time for the mixture reduction algorithms AKL (red), SKL (green) and ISE (blue) is given versus the number of remaining components **M** in the reduced mixture. The original density in the simulations has 25 components.

IV. Summary of the numerical simulation

Three mixture reduction algorithms for 14 commonly used exponential family distributions are compared in Monte-Carlo simulations. The error ensued from the approximation of a mixture density with another density with less components is quantified in terms of integral square error (ISE) between the original mixture and its approximation via mixture reduction algorithms. Similar comparison is performed with respect to the Kullback-Leibler divergence (KLD). Not surprisingly, ISE mixture reduction algorithm performs best when the comparison is with respect to the ISE. Although AKL uses an approximation of KLD between the original mixture and its approximation, it performs best when the comparison is in terms of the KLD with an exception of gamma distribution and inverse gamma distribution. SKL is a heuristic which can be used at a lower computational cost and accuracy.

Although in the overall comparison of SKL with AKL in terms of ISE, AKL is pointed out as the winner, the comparison does not give a uniform result. Indeed, a uniform result is not expected since none of the two algorithms are designed to perform better in terms of ISE. Similarly, in comparison of SKL with ISE in terms of KLD, no algorithm is obviously better than the other one. Again, none of these two algorithms are tailored to obtain lower error with respect to KLD. It is intuitive to believe ISE should perform worse than SKL in terms of KLD (especially at increased reduction aggressiveness), since ISE may select a pruning hypothesis while SKL can not. Pruning may remove the probability mass from regions in the support which have a considerable probability mass. Hence, KLD between the original mixture density and its approximate may tend to infinity. Contrary to this intuition, for those densities whose mixture density in this simulation does not turn out to be multimodal, such as exponential distribution, Laplace distribution with known mean, Rayleigh distribution and inverse gamma distribution, the ISE performs better than SKL with respect to KLD. This observation is due to the fact that in these densities the probability mass is concentrated in a single mode. Thus, the problem of division by (nearly) zero in calculation of KLD after a pruning is not common.

The comparison of the mixture reduction algorithms in terms of their cycle time shows that the most costly algorithm is the ISE and the least costly is the SKL. When the computational budget permits AKL is recommended by the authors since it has a comparable performance to ISE but is computationally less heavy.

REFERENCES

[1] S.S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*, Artech House radar library. Artech House, 1999.

[2] D.L. Alspach and H.W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *Automatic Control, IEEE Transactions on*, vol. 17, no. 4, pp. 439–448, 1972.

[3] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4091–4104, nov. 2006.

[4] D.J. Salmond, "Mixture reduction algorithms for target tracking," in *State Estimation in Aerospace and Tracking Applications, IEE Colloquium on*, London, UK, dec 1989, pp. 7/1–7/4.

[5] D.J. Salmond, "Mixture reduction algorithms for point and extended object tracking in clutter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 45, no. 2, pp. 667–686, april 2009.

[6] O. Bochardt, R. Calhoun, J.K. Uhlmann, and S.J. Julier, "Generalized information representation and compression using covariance union," in *Information Fusion, 2006 9th International Conference on*, Florence, Italy, july 2006, pp. 1–7.

[7] S. Reece and S. Roberts, "Generalised covariance union: A unified approach to hypothesis merging in tracking," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 1, pp. 207–221, jan. 2010.

[8] D.F. Crouse, P. Willett, K. Pattipati, and L. Svensson, "A look at Gaussian mixture reduction algorithms," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, july 2011, pp. 1–8.

[9] M.F. Huber and U.D. Hanebeck, "Progressive Gaussian mixture reduction," in *Information Fusion, 2008 11th International Conference on*, Cologne, Germany, july 2008, pp. 1–8.

[10] D. Schieferdecker and M.F. Huber, "Gaussian mixture reduction via clustering," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, Seattle, WA, USA, july 2009, pp. 1536–1543.

[11] Jørgen E. Harmse, "Reduction of Gaussian mixture models by maximum similarity," *Journal of Nonparametric Statistics*, vol. 22, no. 6, pp. 703–709, 2010.

[12] Pierrick Bruneau, Marc Gelgon, and Fabien Picarougne, "Parsimonious reduction of Gaussian mixture models with a variational-Bayes approach," *Pattern Recogn.*, vol. 43, pp. 850–858, March 2010.

[13] T. Ardeshiri, U. Orguner, C. Lundquist, and T.B. Schon, "On mixture reduction for multiple target tracking," in *Information Fusion (FUSION), 2012 15th International Conference on*, july 2012, pp. 692 –699.

[14] J. L. Williams and P. S. Maybeck, "Cost-function-based hypothesis control techniques for multiple hypothesis tracking," *Mathematical and Computer Modelling*, vol. 43, no. 9-10, pp. 976–989, May 2006.

[15] A.R. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 43, no. 3, pp. 989–999, july 2007.

[16] J. L. Williams, "Gaussian mixture reduction for tracking multiple maneuvering targets in clutter," M.S. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, USA, 2003.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[18] T. Ardeshiri, K. Granström, E. Özkan, and U. Orguner, "Greedy reduction algorithms for mixtures of exponential family," *Submitted to Signal Processing Letters*, 2014.

[19] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection theory*, Prentice Hall signal processing series. Prentice-Hall PTR, 1998.

[20] J.R. Hershey and P.A. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, april 2007, vol. 4, pp. IV–317 –IV–320.

[21] Genshiro Kitagawa, "The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother," *Annals of the Institute of Statistical Mathematics*, vol. 46, no. 4, pp. 605–623, 1994.

[22] X. Chen, R. Tharmarasa, M. Pelletier, and T. Kirubarajan, "Integrated clutter estimation and target tracking using Poisson point processes," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 2, pp. 1210–1235, Apr. 2012.

[23] K. Granström and U. Orguner, "On the reduction of Gaussian inverse Wishart mixtures," in *Information Fusion (FUSION), 2012 15th International Conference on*, july 2012, pp. 2162 –2169.

[24] K. Granström and U. Orguner, "Estimation and maintenance of measurement rates for multiple extended target tracking," in *Information Fusion (FUSION), 2012 15th International Conference on*, july 2012, pp. 2170 –2176.

[25] M.J. Wainwright and M.I. Jordan, *Graphical Models, Exponential Families, and Variational Inference*, Foundations and trends in machine learning. Now Publishers, 2008.

[26] Ralf Herbrich, "Minimising the Kullback-Leibler divergence," http://research.microsoft.com/pubs/74555/KL.pdf, August 2005.

[27] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 44, no. 3, pp. 1042–1059, July 2008.

[28] Jean Dickinson Gibbons and Subhabrata Chakraborti, *Nonparametric Statistical Inference, Fifth Edition*, Statistics, textbooks and monographs. Taylor & Francis, 2010.

APPENDIX

A. Multiple hypothesis testing

Here the summary of proof for the MAP decision rule is given for the sake of completeness of the technical report. For more complete treatment see [19].

Consider that we want to decide among M hypotheses $\{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_M\}$. Let the the cost assigned to the decision to choose $\mathcal{H}_i$ when $\mathcal{H}_j$ is true is denoted by $\mathcal{C}_{ij}$ where

$$\mathcal{C}_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} . \tag{51}$$

The expected Bayes risk [19] becomes

$$\mathcal{R} = \sum_{i=1}^{M} \sum_{j=1}^{M} \mathcal{C}_{ij} P(\mathcal{H}_i|\mathcal{H}_j) P(\mathcal{H}_j). \tag{52}$$

We are looking for a decision rule that minimizes $\mathcal{R}$. Let us partition the space to regions R_i for $i = 1 : M$ so that

$$\begin{aligned} \mathcal{R} &= \sum_{i=1}^{M} \sum_{j=1}^{M} \mathcal{C}_{ij} \int_{R_i} p(x|\mathcal{H}_j) P(\mathcal{H}_j) \, \mathrm{d}x \\ &= \sum_{i=1}^{M} \int_{R_i} \sum_{j=1}^{M} \mathcal{C}_{ij} P(\mathcal{H}_j|x) p(x) \, \mathrm{d}x \\ &= \sum_{i=1}^{M} \int_{R_i} \mathcal{C}_i p(x) \, \mathrm{d}x \end{aligned} \tag{53}$$

where $\mathcal{C}_i(x) = \sum_{j=1}^{M} \mathcal{C}_{ij} P(\mathcal{H}_j|x)$. Since each data x should trigger only one decision, i.e. assigned to only one of the R_i partitions we should decide $\mathcal{H}_k$ for which $\mathcal{C}_i$ is minimum.

Since $\mathcal{C}_i(x) = \sum_{j=1}^{M} P(\mathcal{H}_j|x) - P(\mathcal{H}_i|x)$, $\mathcal{C}_i(x)$ is minimized if $P(\mathcal{H}_i|x)$ is maximized. Thus the decision rule is decide $\mathcal{H}_k$ if $P(\mathcal{H}_k|x) > P(\mathcal{H}_i|x)$ for $i \neq k$. For equal prior probabilities $P(\mathcal{H}_k) = P(\mathcal{H}_i)$ the decision rule will be to decide $\mathcal{H}_k$ if $p(x|\mathcal{H}_k) > p(x|\mathcal{H}_i)$ for $i \neq k$. This decision rule is also referred to as maximum a posteriori decision rule.

If the prior probabilities are not equal due to e.g., heuristics $P(\mathcal{H}_k) \neq P(\mathcal{H}_i)$, Bayes rule $P(\mathcal{H}_i|x) \propto p(x|\mathcal{H}_i) P(\mathcal{H}_i)$ can be used. This possibility is not exploited in this report.

B. Essential expressions for the reduction of mixtures of exponential family

Essential expressions and formula for reduction of mixture densities of common exponential family distributions are given in this section. Some functions which are used in the expressions such as the gamma function are defined here for completeness. The gamma function is defined by

$$\Gamma(t) = \int_0^{\infty} x^{t-1} \exp(-x) \, \mathrm{d}x. \tag{54}$$

The multivariate gamma function which is a generalization of the gamma function is

$$\begin{aligned} \Gamma_d(t) &= \int_{S \succ 0} \exp(-\operatorname{Tr}(S)) |S|^{t - \frac{d+1}{2}} \, \mathrm{d}S \\ &= \pi^{d(d-1)/4} \prod_{j=1}^{d} \Gamma\left(t + \frac{1-j}{2}\right). \end{aligned} \tag{55}$$

The digamma function is given as

$$\psi(t) = \frac{d}{dt} \log \Gamma(t) = \frac{\Gamma^{'}(t)}{\Gamma(t)}. \tag{56}$$

The multivariate polygamma function of order n is defined as

$$\psi_d^{(n)}(t) = \frac{d^{n+1}}{dt^{n+1}} \log \Gamma_d(t) = \sum_{j=1}^{d} \frac{d^{n+1}}{dt^{n+1}} \log \Gamma\left(t + \frac{1-j}{2}\right) = \sum_{j=1}^{d} \psi^{(n)}\left(t + \frac{1-j}{2}\right). \tag{57}$$

The multinomial beta function in terms of the gamma function is given as

$$\mathrm{B}_K(\alpha) = \frac{\prod_{j=1}^{K} \Gamma(\alpha_j)}{\Gamma\left(\sum_{j=1}^{K} \alpha_j\right)}. \tag{58}$$

1) Exponential Distribution:

$$\mathrm{Exp}(x;\lambda) = \lambda \exp(-\lambda x) \tag{59a}$$
$$\text{Support: } x \in [0, +\infty) \tag{59b}$$
$$\text{Parameter space: } \lambda \in (0, +\infty) \tag{59c}$$
$$\eta = -\lambda \tag{59d}$$
$$A(\eta) = -\log(-\eta) \tag{59e}$$
$$\nabla_\eta A = \frac{\partial A}{\partial \eta} = -\frac{1}{\eta} \tag{59f}$$
$$h(x) = 1 \tag{59g}$$
$$\mathbb{E}[h(x)] = 1 \tag{59h}$$
$$T(x) = x \tag{59i}$$

Solution to $\nabla_{\eta^L} A = Y$ is given by

$$\eta^L = -\frac{1}{Y}. \tag{60}$$

2) Weibull Distribution with known shape k:

$$\text{Weibull}(x;\lambda,k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp\left(-\frac{x^k}{\lambda^k}\right) \tag{61a}$$

$$\text{Support: } x \in [0,+\infty) \tag{61b}$$

$$\text{Parameter space: } \lambda \in (0,+\infty),\ k \in (0,+\infty) \tag{61c}$$

$$\eta = -\frac{1}{\lambda^k} \tag{61d}$$

$$A(\eta) = -\log(-\eta) - \log(k) \tag{61e}$$

$$\nabla_\eta A = \frac{\partial A}{\partial \eta} = -\frac{1}{\eta} \tag{61f}$$

$$h(x) = x^{k-1} \tag{61g}$$

$$\mathbb{E}[h(x)] = \Gamma\left(\frac{2k-1}{k}\right)(-\eta)^{\frac{1-k}{k}} \tag{61h}$$

$$T(x) = x^k \tag{61i}$$

Solution to $\nabla_{\eta^L} A = Y$ is given by

$$\eta^L = -\frac{1}{Y}. \tag{62}$$

The expression for $\mathbb{E}_{q(x;\eta)}[h(x)]$ is derived here

$$\begin{aligned}\mathbb{E}[h(x)] &= \int_0^\infty x^{k-1}\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp\left(-\frac{x^k}{\lambda^k}\right)\,\mathrm{d}x = \left\{z \triangleq \frac{x^k}{\lambda^k},\ \mathrm{d}z = k\frac{x^{k-1}}{\lambda^k}\,\mathrm{d}x\right\} = \int_0^\infty \lambda^{k-1} z^{\frac{k-1}{k}}\exp(-z)\,\mathrm{d}z \\ &= \lambda^{k-1}\Gamma\left(\frac{k-1}{k}+1\right) = \lambda^{k-1}\Gamma\left(\frac{2k-1}{k}\right) = \Gamma\left(\frac{2k-1}{k}\right)(-\eta)^{\frac{1-k}{k}}.\end{aligned} \tag{63}$$

3) Laplace Distribution with known mean μ:

$$\text{Laplace}(x;\mu,b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) \tag{64a}$$

$$\text{Support: } x \in (-\infty,\infty) \tag{64b}$$

$$\text{Parameter space: } b \in (0,+\infty),\ \mu \in \mathbb{R} \tag{64c}$$

$$\eta = -\frac{1}{b} \tag{64d}$$

$$A(\eta) = \log\left(-\frac{2}{\eta}\right) \tag{64e}$$

$$\nabla_\eta A = \frac{\partial A}{\partial \eta} = -\frac{1}{\eta} \tag{64f}$$

$$h(x) = 1 \tag{64g}$$

$$\mathbb{E}[h(x)] = 1 \tag{64h}$$

$$T(x) = |x-\mu| \tag{64i}$$

Solution to $\nabla_{\eta^L} A = Y$ is given by

$$\eta^L = -\frac{1}{Y}. \tag{65}$$

4) Rayleigh Distribution :

$$\text{Rayleigh}(x;\sigma) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{66a}$$

$$\text{Support: } x \in [0,+\infty) \tag{66b}$$

$$\text{Parameter space: } \sigma \in (0,+\infty) \tag{66c}$$

$$\eta = -\frac{1}{2\sigma^2} \tag{66d}$$

$$A(\eta) = -\log(-2\eta) \tag{66e}$$

$$\nabla_\eta A = \frac{\partial A}{\partial \eta} = -\frac{1}{\eta} \tag{66f}$$

$$h(x) = x \tag{66g}$$

$$\mathbb{E}[h(x)] = \sqrt{\frac{\pi}{-2\eta}} \tag{66h}$$

$$T(x) = x^2 \tag{66i}$$

Solution to $\nabla_{\eta^L} A = Y$ is given by

$$\eta^L = -\frac{1}{Y}. \tag{67}$$

The expression for $\mathbb{E}_{q(x;\eta)}[h(x)]$ is derived here

$$\mathbb{E}[h(x)] = \int_0^\infty x\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)\,\mathrm{d}x = \sigma\sqrt{\frac{\pi}{2}} = \sqrt{\frac{\pi}{-2\eta}}. \tag{68}$$

5) Log-normal Distribution:

$$\log -\mathcal{N}(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(\log x-\mu)^2\right) \tag{69a}$$

$$\text{Support: } x \in (0,+\infty) \tag{69b}$$

$$\text{Parameter space: } \sigma \in (0,+\infty),\ \mu \in \mathbb{R} \tag{69c}$$

$$\eta = (\eta_1,\eta_2) \tag{69d}$$

$$\eta_1 = \frac{\mu}{\sigma^2} \tag{69e}$$

$$\eta_2 = -\frac{1}{2\sigma^2} \tag{69f}$$

$$A(\eta) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \tag{69g}$$

$$\nabla_\eta A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}\right) \tag{69h}$$

$$\frac{\partial A}{\partial \eta_1} = -\frac{\eta_1}{2\eta_2} \tag{69i}$$

$$\frac{\partial A}{\partial \eta_2} = \frac{\eta_1^2}{4\eta_2^2} - \frac{1}{2\eta_2} \tag{69j}$$

$$h(x) = \frac{1}{x\sqrt{2\pi}} \tag{69k}$$

$$\mathbb{E}[h(x)] = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{\eta_1}{2\eta_2} - \frac{1}{4\eta_2}\right) \tag{69l}$$

$$T(x) = \Big(\log(x), (\log(x))^2\Big) \tag{69m}$$

Solution to the system of equations $\nabla_{\eta^L} A = Y$ is given by

$$\eta_2^L = \frac{-2}{Y_2 - Y_1^2}, \tag{70a}$$

$$\eta_1^L = -2Y_1\eta_2^L. \tag{70b}$$

The expression for $\mathbb{E}_{q(x;\eta)}[h(x)]$ is derived here

$$\mathbb{E}[h(x)] = \frac{1}{\sqrt{2\pi}}\,\mathbb{E}\left[\frac{1}{x}\right] = \frac{1}{\sqrt{2\pi}}\exp\left(-\mu + \frac{\sigma^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{\eta_1}{2\eta_2} - \frac{1}{4\eta_2}\right). \tag{71}$$

6) Gamma Distribution:

$$\mathrm{Gamma}(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x) \tag{72a}$$

$$\text{Support: } x \in (0,+\infty) \tag{72b}$$

$$\text{Parameter space: } \alpha \in (0,+\infty),\ \beta \in (0,+\infty) \tag{72c}$$

$$\eta = (\eta_1, \eta_2) \tag{72d}$$

$$\eta_1 = \alpha - 1 \tag{72e}$$

$$\eta_2 = -\beta \tag{72f}$$

$$A(\eta) = \log\Gamma(\eta_1+1) - (\eta_1+1)\log(-\eta_2) \tag{72g}$$

$$\nabla_\eta A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}\right) \tag{72h}$$

$$\frac{\partial A}{\partial \eta_1} = \psi(\eta_1+1) - \log(-\eta_2) \tag{72i}$$

$$\frac{\partial A}{\partial \eta_2} = -\frac{\eta_1+1}{\eta_2} \tag{72j}$$

$$h(x) = 1 \tag{72k}$$

$$\mathbb{E}[h(x)] = 1 \tag{72l}$$

$$T(x) = (\log(x), x) \tag{72m}$$

To solve the system of equations $\nabla_{\eta^L} A = Y$, first let $Z = \log(Y_2) - Y_1$ and $u = \eta_1 + 1$. Then solve $\psi(u) - \log(u) + Z = 0$ numerically and obtain

$$\eta_1^L = u - 1, \tag{73a}$$

$$\eta_2^L = -\frac{u}{Y_2}. \tag{73b}$$

7) Inverse Gamma Distribution:

$$\mathrm{IGamma}(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) \tag{74a}$$

$$\text{Support: } x \in (0,+\infty) \tag{74b}$$

$$\text{Parameter space: } \alpha \in (0,+\infty),\ \beta \in (0,+\infty) \tag{74c}$$

$$\eta = (\eta_1, \eta_2) \tag{74d}$$

$$\eta_1 = -\alpha - 1 \tag{74e}$$

$$\eta_2 = -\beta \tag{74f}$$

$$A(\eta) = \log\Gamma(-\eta_1 - 1) - (-\eta_1 - 1)\log(-\eta_2) \tag{74g}$$

$$\nabla_{\eta} A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}\right) \tag{74h}$$

$$\frac{\partial A}{\partial \eta_1} = -\psi(-\eta_1 - 1) + \log(-\eta_2) \tag{74i}$$

$$\frac{\partial A}{\partial \eta_2} = \frac{\eta_1 + 1}{\eta_2} \tag{74j}$$

$$h(x) = 1 \tag{74k}$$

$$\mathbb{E}[h(x)] = 1 \tag{74l}$$

$$T(x) = \left(\log(x), \frac{1}{x}\right) \tag{74m}$$

To solve the system of equations $\nabla_{\eta^L} A = Y$ first let $Z = \log(Y_2) + Y_1$ and $u = -\eta_1 - 1$. Then solve $\psi(u) + \log(u) - Z = 0$ numerically and obtain

$$\eta_1^L = -u - 1, \tag{75a}$$

$$\eta_2^L = -\frac{u}{Y_2}. \tag{75b}$$

8) Univariate Gaussian Distribution:

$$\mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \tag{76a}$$

$$\text{Support: } x \in \mathbb{R} \tag{76b}$$

$$\text{Parameter space: } \sigma \in (0,+\infty),\ \mu \in \mathbb{R} \tag{76c}$$

$$\eta = (\eta_1, \eta_2) \tag{76d}$$

$$\eta_1 = \frac{\mu}{\sigma^2} \tag{76e}$$

$$\eta_2 = -\frac{1}{2\sigma^2} \tag{76f}$$

$$A(\eta) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \tag{76g}$$

$$\nabla_\eta A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}\right) \tag{76h}$$

$$\frac{\partial A}{\partial \eta_1} = -\frac{\eta_1}{2\eta_2} \tag{76i}$$

$$\frac{\partial A}{\partial \eta_2} = \frac{\eta_1^2}{4\eta_2^2} - \frac{1}{2\eta_2} \tag{76j}$$

$$h(x) = \frac{1}{\sqrt{2\pi}} \tag{76k}$$

$$\mathbb{E}[h(x)] = \frac{1}{\sqrt{2\pi}} \tag{76l}$$

$$T(x) = \left(x, x^2\right) \tag{76m}$$

Solution to the system of equations $\nabla_{\eta^L} A = Y$ is given by

$$\eta_2^L = \frac{-2}{Y_2 - Y_1^2}, \tag{77a}$$

$$\eta_1^L = -2Y_1\eta_2^L. \tag{77b}$$

9) Multivariate Gaussian Distribution:

$$\mathcal{N}(x; m, P) = (2\pi)^{-\frac{k}{2}} |P|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-m)^{\mathrm{T}} P^{-1}(x-m)\right) \tag{78a}$$

$$\text{Support: } x \in \mathbb{R}^k \tag{78b}$$

$$\text{Parameter space: } P \in \mathbb{R}^{k\times k} \text{ and } P = P^{\mathrm{T}} \succ 0,\ \mu \in \mathbb{R}^k \tag{78c}$$

$$\eta = (\eta_1, \eta_2) \tag{78d}$$

$$\eta_1 = P^{-1} m \tag{78e}$$

$$\eta_2 = -\frac{1}{2} P^{-1} \tag{78f}$$

$$A(\eta) = -\frac{1}{4}\eta_1^{\mathrm{T}} \eta_2^{-1} \eta_1 - \frac{1}{2}\log|-2\eta_2| \tag{78g}$$

$$\nabla_\eta A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}\right) \tag{78h}$$

$$\frac{\partial A}{\partial \eta_1} = -\frac{1}{2}\eta_1^{\mathrm{T}} \eta_2^{-1} \tag{78i}$$

$$\frac{\partial A}{\partial \eta_2} = \frac{1}{4}\eta_2^{-\mathrm{T}} \eta_1 \eta_1^{\mathrm{T}} \eta_2^{-\mathrm{T}} - \frac{1}{2}\eta_2^{-1} \tag{78j}$$

$$h(x) = (2\pi)^{-k/2} \tag{78k}$$

$$\mathbb{E}[h(x)] = (2\pi)^{-k/2} \tag{78l}$$

$$T(x) = \left(x, xx^{\mathrm{T}}\right) \tag{78m}$$

Solution to system of equations $\nabla_{\eta^L} A = Y$ is given by

$$\eta_2^L = -\frac{1}{2}(Y_2 - Y_1^{\mathrm{T}} Y_1)^{-1}, \tag{79}$$

$$\eta_1^L = \left(-2Y_1 \eta_2^L\right)^{\mathrm{T}}. \tag{80}$$

10) Gaussian Gamma Distribution:

$$\text{GaussianGamma}\left(x,\tau;\mu,\lambda,\alpha,\beta\right)=\mathcal{N}\left(x;\mu,\frac{1}{\lambda\tau}\right)\text{Gamma}(\tau;\alpha,\beta) \tag{81a}$$

$$\text{Support: } x\in\mathbb{R},\ \tau\in(0,+\infty) \tag{81b}$$

$$\text{Parameter space: } \alpha\in(0,+\infty),\ \beta\in(0,+\infty),\ \lambda\in(0,+\infty),\ \mu\in\mathbb{R} \tag{81c}$$

$$\eta=(\eta_1,\eta_2,\eta_3,\eta_4) \tag{81d}$$

$$\eta_1=\alpha-\frac{1}{2} \tag{81e}$$

$$\eta_2=-\beta-\frac{\lambda\mu^2}{2} \tag{81f}$$

$$\eta_3=\lambda\mu \tag{81g}$$

$$\eta_4=-\frac{\lambda}{2} \tag{81h}$$

$$A(\eta)=\log\Gamma\left(\eta_1+\frac{1}{2}\right)-\frac{1}{2}\log(-2\eta_4)-\left(\eta_1+\frac{1}{2}\right)\log\left(-\eta_2+\frac{\eta_3^2}{4\eta_4}\right) \tag{81i}$$

$$\nabla_\eta A=\left(\frac{\partial A}{\partial\eta_1},\frac{\partial A}{\partial\eta_2},\frac{\partial A}{\partial\eta_3},\frac{\partial A}{\partial\eta_4}\right) \tag{81j}$$

$$\frac{\partial A}{\partial\eta_1}=\psi\left(\eta_1+\frac{1}{2}\right)-\log\left(-\eta_2+\frac{\eta_3^2}{4\eta_4}\right) \tag{81k}$$

$$\frac{\partial A}{\partial\eta_2}=-\frac{\eta_1+\frac{1}{2}}{-\eta_2+\frac{\eta_3^2}{4\eta_4}} \tag{81l}$$

$$\frac{\partial A}{\partial\eta_3}=-\frac{2\eta_3+\left(\eta_1+\frac{1}{2}\right)}{4\eta_4\left(-\eta_2+\frac{\eta_3^2}{4\eta_4}\right)} \tag{81m}$$

$$\frac{\partial A}{\partial\eta_4}=\frac{\eta_3^2\left(\eta_1+\frac{1}{2}\right)}{4\eta_4^2\left(-\eta_2+\frac{\eta_3^2}{4\eta_4}\right)}-\frac{1}{2\eta_4} \tag{81n}$$

$$h(x)=\frac{1}{\sqrt{2\pi}} \tag{81o}$$

$$\mathbb{E}[h(x)]=\frac{1}{\sqrt{2\pi}} \tag{81p}$$

$$T(x)=\left(\log(\tau),\tau,\tau x,\tau x^2\right) \tag{81q}$$

To solve the system of equations $\nabla_{\eta^L}A=Y$ first let $Z=\log(-Y_2)-Y_1$ and $u=\eta_1+\frac{1}{2}$. Then solve $\psi(u)-\log(u)+Z=0$ numerically and obtain

$$\eta_1^L=u-\frac{1}{2}, \tag{82a}$$

$$\eta_4^L=-\frac{1}{2}\left(\frac{Y_3^2}{Y_2}+Y_4\right)^{-1}, \tag{82b}$$

$$\eta_3^L=\frac{2\eta_4 Y_3}{Y_2}, \tag{82c}$$

$$\eta_2^L=\frac{\eta_3^2}{4\eta_4}+\frac{\eta_1+\frac{1}{2}}{Y_2}. \tag{82d}$$

11) Dirichlet distribution:

$$\mathrm{Dir}_K(x;\alpha) = \frac{1}{\mathrm{B}(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1} \tag{83a}$$

$$\text{Support: } x_i \in [0,1] \text{ for } i = 1 \cdots K \text{ and } \sum_{i=1}^{K} x_i = 1 \tag{83b}$$

$$\text{Parameter space: } \alpha_i > 0 \text{ and } K \geq 2 \tag{83c}$$

$$\eta = (\eta_1, \cdots, \eta_K) \tag{83d}$$

$$\eta_i = \alpha_i - 1 \tag{83e}$$

$$A(\eta) = \sum_{i=1}^{K} \log \Gamma(\eta_i + 1) - \log \Gamma \left(\sum_{i=1}^{K} (\eta_i + 1) \right) \tag{83f}$$

$$\nabla_\eta A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}, \cdots, \frac{\partial A}{\partial \eta_K} \right) \tag{83g}$$

$$\frac{\partial A}{\partial \eta_i} = \psi(\eta_i + 1) - \psi \left(\sum_{i=1}^{K} (\eta_i + 1) \right) \tag{83h}$$

$$h(x) = 1 \tag{83i}$$

$$\mathbb{E}[h(x)] = 1 \tag{83j}$$

$$T(x) = (\log(x_1), \cdots, \log(x_K)) \tag{83k}$$

The system of equations $\nabla_{\eta^L} A = Y$ can be solved using a numerical method such as newton method where the Hessian is given by,

$$\frac{\partial^2 A}{\partial {\eta_i}^2} = \psi^{(1)}(\eta_i + 1) - \psi^{(1)} \left(\sum_{k=1}^{K} (\eta_k + 1) \right), \tag{84a}$$

$$\frac{\partial^2 A}{\partial \eta_{ij}} = -\psi^{(1)} \left(\sum_{k=1}^{K} (\eta_k + 1) \right). \tag{84b}$$

12) Wishart Distribution:

$$\mathcal{W}_d\left(X;n,V\right) = \frac{|X|^{\frac{1}{2}(n-d-1)} \exp \operatorname{Tr}\left(-\frac{1}{2}V^{-1}X\right)}{2^{\frac{1}{2}nd}\Gamma_d\left(\frac{1}{2}n\right)|V|^{\frac{1}{2}n}} \tag{85a}$$

$$\text{Support:} X \in \mathbb{R}^{d\times d} \text{ and } X = X^{\mathrm{T}} \succ 0 \tag{85b}$$

$$\text{Parameter space: } V \in \mathbb{R}^{d\times d} \text{ and } V = V^{\mathrm{T}} \succ 0,\ n \geq d \tag{85c}$$

$$\eta = (\eta_1, \eta_2) \tag{85d}$$

$$\eta_1 = \frac{1}{2}(n-d-1) \tag{85e}$$

$$\eta_2 = -\frac{1}{2}V^{-1} \tag{85f}$$

$$A(\eta) = -\left(\eta_1 + \frac{d+1}{2}\right)\log|-\eta_2| + \log\Gamma_d\left(\eta_1 + \frac{d+1}{2}\right) \tag{85g}$$

$$\nabla_\eta A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}\right) \tag{85h}$$

$$\frac{\partial A}{\partial \eta_1} = -\log|-\eta_2| + \psi_d\left(\eta_1 + \frac{d+1}{2}\right) \tag{85i}$$

$$\frac{\partial A}{\partial \eta_2} = -\left(\eta_1 + \frac{d+1}{2}\right)\eta_2^{-1} \tag{85j}$$

$$h(x) = 1 \tag{85k}$$

$$\mathbb{E}[h(x)] = 1 \tag{85l}$$

$$T(x) = (\log|X|, X) \tag{85m}$$

To solve the system of equations $\nabla_{\eta^L} A = Y$ first let $Z = \log|Y_2| - Y_1$ and $u = \eta_1 + \frac{d+1}{2}$. Then solve $\psi_d(u) - d\log(u) + Z = 0$ numerically and obtain

$$\eta_1^L = u - \frac{d+1}{2}, \tag{86a}$$

$$\eta_2^L = -uY_2^{-1}. \tag{86b}$$

13) Inverse Wishart Distribution:

$$\mathcal{IW}_d\left(X;\nu,\Psi\right) = \frac{|\Psi|^{\frac{1}{2}(\nu-d-1)}\exp\operatorname{Tr}\left(-\frac{1}{2}\Psi X^{-1}\right)}{2^{\frac{1}{2}(\nu-d-1)d}\Gamma_d\left(\frac{1}{2}(\nu-d-1)\right)|X|^{\frac{1}{2}\nu}} \tag{87a}$$

$$\text{Support: } X\in\mathbb{R}^{d\times d} \text{ and } X=X^{\mathrm{T}}\succ 0 \tag{87b}$$

$$\text{Parameter space: } \nu>2d\ \Psi\in\mathbb{R}^{d\times d},\ \Psi=\Psi^{\mathrm{T}}\succ 0 \tag{87c}$$

$$\eta=(\eta_1,\eta_2) \tag{87d}$$

$$\eta_1=-\frac{1}{2}\nu \tag{87e}$$

$$\eta_2=-\frac{1}{2}\Psi \tag{87f}$$

$$A(\eta)=\left(\eta_1+\frac{d+1}{2}\right)\log|-\eta_2|+\log\Gamma_d\left(-\eta_1-\frac{d+1}{2}\right) \tag{87g}$$

$$\nabla_\eta A=\left(\frac{\partial A}{\partial\eta_1},\frac{\partial A}{\partial\eta_2}\right) \tag{87h}$$

$$\frac{\partial A}{\partial\eta_1}=\log|-\eta_2|-\psi_d\left(-\eta_1-\frac{d+1}{2}\right) \tag{87i}$$

$$\frac{\partial A}{\partial\eta_2}=\left(\eta_1+\frac{d+1}{2}\right)\eta_2^{-1} \tag{87j}$$

$$h(x)=1 \tag{87k}$$

$$\mathbb{E}[h(x)]=1 \tag{87l}$$

$$T(x)=\left(\log|X|,X^{-1}\right) \tag{87m}$$

To solve the system of equations $\nabla_{\eta^L}A=Y$ first let $Z=-\log(Y_2)-Y_1$ and $u=-\eta_1-\frac{d+1}{2}$. Then solve $-\psi_d(u)+d\log(u)+Z=0$ numerically and obtain

$$\eta_1^L=-u-\frac{d+1}{2}, \tag{88a}$$

$$\eta_2^L=-uY_2^{-1}. \tag{88b}$$

14) Gaussian Inverse Wishart Distribution:

$$\mathrm{GIW}\left(x, X; m, P, \nu, \Psi\right) = \mathcal{N}\left(x; m, P\right) \mathcal{IW}_d\left(X; \nu, \Psi\right) \tag{89a}$$

$$\text{Support: } x \in \mathbb{R}^k,\ X \in \mathbb{R}^{d\times d} \text{ and } X = X^{\mathrm{T}} \succ 0 \tag{89b}$$

$$\text{Parameter space: } \nu > 2d\ \Psi \in \mathbb{R}^{d\times d},\ \Psi = \Psi^{\mathrm{T}} \succ 0,\ P \in \mathbb{R}^{k\times k} \text{ and } P = P^{\mathrm{T}} \succ 0,\ \mu \in \mathbb{R}^k \tag{89c}$$

$$\eta = (\eta_1, \eta_2, \eta_3, \eta_4) \tag{89d}$$

$$\eta_1 = -\frac{1}{2}\nu \tag{89e}$$

$$\eta_2 = -\frac{1}{2}\Psi \tag{89f}$$

$$\eta_3 = P^{-1}m \tag{89g}$$

$$\eta_4 = -\frac{1}{2}P^{-1} \tag{89h}$$

$$A(\eta) = \left(\eta_1 + \frac{d+1}{2}\right)\log|-\eta_2| + \log\Gamma_d\left(-\eta_1 - \frac{d+1}{2}\right) - \frac{1}{4}\eta_3^{\mathrm{T}}\eta_4^{-1}\eta_3 - \frac{1}{2}\log|-2\eta_4| \tag{89i}$$

$$\nabla_\eta A = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}, \frac{\partial A}{\partial \eta_3}, \frac{\partial A}{\partial \eta_4}\right) \tag{89j}$$

$$\frac{\partial A}{\partial \eta_1} = \log|-\eta_2| - \psi_d\left(-\eta_1 - \frac{d+1}{2}\right) \tag{89k}$$

$$\frac{\partial A}{\partial \eta_2} = \left(\eta_1 + \frac{d+1}{2}\right)\eta_2^{-1} \tag{89l}$$

$$\frac{\partial A}{\partial \eta_3} = -\frac{1}{2}\eta_3^{\mathrm{T}}\eta_4^{-1} \tag{89m}$$

$$\frac{\partial A}{\partial \eta_4} = \frac{1}{4}\eta_4^{-\mathrm{T}}\eta_1\eta_3^{\mathrm{T}}\eta_4^{-\mathrm{T}} - \frac{1}{2}\eta_4^{-1} \tag{89n}$$

$$h(x) = (2\pi)^{-k/2} \tag{89o}$$

$$\mathbb{E}[h(x)] = (2\pi)^{-k/2} \tag{89p}$$

$$T(x) = \left(\log|X|, X^{-1}, x, xx^{\mathrm{T}}\right) \tag{89q}$$

To solve the system of equations $\nabla_{\eta^L} A = Y$ first let $Z = -\log(Y_2) - Y_1$ and $u = -\eta_1 - \frac{d+1}{2}$. Then solve $-\psi_d(u) + d\log(u) + Z = 0$ numerically and obtain

$$\eta_1^L = -u - \frac{d+1}{2}, \tag{90a}$$

$$\eta_2^L = -uY_2^{-1}, \tag{90b}$$

$$\eta_3^L = -\frac{1}{2}(Y_4 - Y_3^{\mathrm{T}}Y_3)^{-1}, \tag{90c}$$

$$\eta_4^L = \left(-2Y_3\eta_4^L\right)^{\mathrm{T}}. \tag{90d}$$

Avdelning, Institution
Division, Department

Division of Automatic Control
Department of Electrical Engineering

Datum
Date

2014-08-18

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English

☐ _____

Rapporttyp
Report category

☐ Licentiatavhandling
☐ Examensarbete
☐ C-uppsats
☐ D-uppsats
☒ Övrig rapport
☐ _____

ISBN
—

ISRN
—

Serietitel och serienummer
Title of series, numbering

ISSN
1400-3902

LiTH-ISY-R-3076

URL för elektronisk version

`http://www.control.isy.liu.se`

Titel
Title

On Reduction of Mixtures of the Exponential Family Distributions

Författare
Author

Tohid Ardeshiri, Emre Özkan , Umut Orguner

Sammanfattning
Abstract

Many estimation problems require a mixture reduction algorithm with which an increasing number of mixture components are reduced to a tractable level. In this technical report a discussion on different aspects of mixture reduction is given followed by a presentation of numerical simulation on reduction of mixture densities where the component density belongs to the exponential family of distributions.

Nyckelord
Keywords

Mixture density, mixture reduction, exponential family, integral square error, Kullback-Leibler divergence, Exponential Distribution, Weibull Distribution, Laplace Distribution, Rayleigh Distribution, Log-normal Distribution, Gamma Distribution, Inverse Gamma Distribution, Univariate Gaussian Distribution, Multivariate Gaussian Distribution, Gaussian Gamma Distribution, Dirichlet distribution, Wishart Distribution, Inverse Wishart Distribution, Gaussian Inverse Wishart Distribution.