Gaussian Mixture Based Progressive Chernoff Fusion

Simone Semeraro, Keith A. LeGrand
Sensing, Controls, and Probabilistic Estimation (SCOPE) Group
School of Aeronautics and Astronautics
Purdue University
West Lafayette, IN 47907, United States
ssemerar@purdue.edu, klegrand@purdue.edu

***Abstract*—Probabilistic decentralized data fusion is the process of combining probabilistic beliefs from multiple sensors to reduce uncertainty and is broadly applicable to problems in aerospace, robotics, and wireless sensor networks. Fusing statistical information into a fused density is especially useful in distributed sensor networks, where each node or agent possesses only limited computation and sensing capabilities. The Chernoff fusion rule prevents information double-counting and produces a fused density that is equidistant from the input densities in an information-theoretic sense. This paper presents a novel approach to Gaussian mixture Chernoff fusion based on the progressive Bayes framework, where the optimal fused mixture is obtained through homotopy continuation. An advantage of the new approach is that it does not require fitting of Gaussian mixtures to lossy samples. A new and simple strategy for determining the optimal Chernoff weighting parameter is also presented and shown to outperform more complicated methods.**

***Index Terms*—Ad hoc networks, Bayesian methods, exponential mixture density fusion, Chernoff fusion, Gaussian mixture approximation, Monte Carlo methods, non-Gaussian fusion, optimization, sensor fusion.**

I. Introduction

In state estimation problems, distributed data fusion enables the dissemination of local information across a sensor network to collectively improve the network agents' estimates. Information can be disseminated in the form of measurements or statistics, the latter of which is generally advantageous in large and/or communication-limited networks [1], [2]. While Bayes-optimal fusion of statistics, such as probability density functions (pdfs), is possible, the associated level of bookkeeping and knowledge of network topology required is prohibitive, especially in *ad hoc* networks [3]. On the other hand, suboptimal *conservative* approaches, such as algebraic mean density (AMD) and exponential mixture density (EMD) fusion, require no knowledge of the network topology [4], [5]. The EMD fusion rule is known under numerous names, including geometric mean density (GMD), weighted exponential product (WEP), and generalized covariance intersection (GCI), to name a few.

The Chernoff fusion rule is a special case of EMD fusion. Chernoff fusion is theoretically appealing as the resulting fused density is equidistant from the input densities in a Kullback-Leibler divergence sense [6], [7]. Because no closed-form solution of the Chernoff fusion rule (and more generally, EMD) is available for Gaussian mixture (GM) input pdfs, existing work has focused on developing fast and accurate GM approximations. Work in [8], [9] fits a GM approximation to weighted stochastic samples drawn from a first-order approximate proposal density. An alternative approach is offered in [10], [11], where GMs are fit to the exponentiated input pdfs using a sigma point approximation. Despite these recent advances, no existing methods for GM Chernoff fusion offer a systematic mechanism for constraining the approximation error to an arbitrary tolerance.

This paper presents the Progressive Chernoff fusion algorithm, a novel fusion algorithm inspired by Hanebeck's Progressive Bayes framework [12] for nonlinear estimation. As illustrated in Fig. 1, the Progressive Chernoff algorithm employs a homotopy continuation to gradually adjust the solution of a simplified problem to the more complicated target solution. A key feature of this approach is that it dynamically adjusts the output mixture size to satisfy any arbitrary error tolerance (for a given fusion parameter), and thus is highly flexible. This paper also considers the problem of selecting the Chernoff optimal fusion weight and proposes a simple yet accurate optimization method.

II. Problem Formulation

Consider two univariate pdfs $p_1(x)$ and $p_2(x)$ that are known and represented as the GMs

$$\begin{aligned} p_1(x) &= \sum_{r=1}^{M} w_{1,r}\mathcal{N}(x;\mu_{1,r},\sigma_{1,r}^2) \\ p_2(x) &= \sum_{l=1}^{N} w_{2,l}\mathcal{N}(x;\mu_{2,l},\sigma_{2,l}^2) \end{aligned} \tag{1}$$

where $w_{(\cdot)}$ are the GM weights and

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

denotes a normal density over the variable x with mean μ and standard deviation σ. The density $p_1(x)$ is assumed to be absolutely continuous with respect to $p_2(x)$. In the context

The authors would like to thank the Boeing Company's Enterprise Technology Office for their support under contract SSOW-BRT-Z0722-5045.

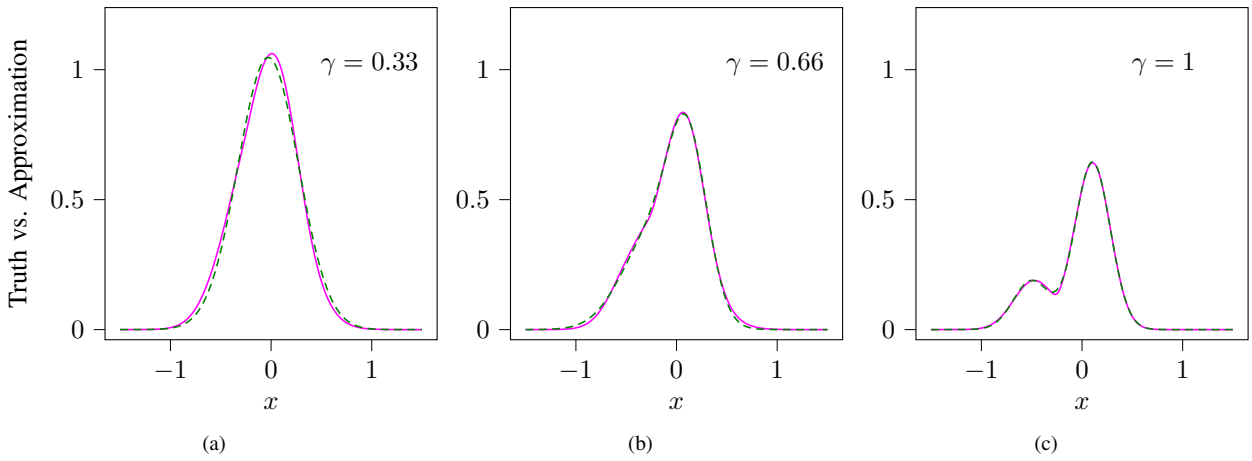

Fig. 1. An example of Progressive Chernoff fusion, where the solution parameters are continuously adjusted from a known simple result to the target density. Three intermediate solutions of the homotopy continuation are pictured, where the true target and approximation are represented by the solid magenta dashed green curves, respectively.

of distributed data fusion, the densities $p_1(x)$ and $p_2(x)$ may represent the probabilistic estimates of two sensing agents conditioned on different measurement data. The underlying data from which $p_1(x)$ and $p_2(x)$ were formed is unknown to the fusion agent. Thus, this paper considers the general family of EMD [3] fusion rules, which are robust to rumor propagation. The fused EMD density is defined as

$$f_{\text{EMD}}(x) = \frac{p_1(x)^{\omega} p_2(x)^{1-\omega}}{\int_{\mathbb{R}} p_1(x)^{\omega} p_2(x)^{1-\omega} dx} \tag{3}$$

where the fusion weight parameter $\omega \in [0,1]$ controls the degree to which each agent's input contributes to the fused solution.

A special case of EMD fusion is Chernoff fusion [6], [13], defined by

$$f_C(x) = \frac{p_1(x)^{\omega^*} p_2(x)^{1-\omega^*}}{\int_{\mathbb{R}} p_1(x)^{\omega^*} p_2(x)^{1-\omega^*} dx} \tag{4}$$

$$\omega^* = \arg \min_{\omega \in [0,1]} \int_{\mathbb{R}} p_1(x)^{\omega} p_2(x)^{1-\omega} dx \tag{5}$$

where the optimal fusion parameter ω^* minimizes the Chernoff information. By this rule, the information gained from each input is equal. Note that this property may be less desirable when stark differences in informativeness exist between sensors, in which case less-informed sensors can "dilute" the fused density and slow convergence in consensus-based architectures. In such circumstances, other weight selection rules may achieve faster consensus convergence [14], [15]. For the purpose of evaluating the suboptimality of various solution methods, the associated optimal cost

$$J^* = \min_{\omega \in [0,1]} \int_{\mathbb{R}} p_1(x)^{\omega} p_2(x)^{1-\omega} dx = \min_{\omega \in [0,1]} J_{\omega}(\omega) \tag{6}$$

is defined.

This paper addresses the problem of finding a GM approximation of the optimal Chernoff fused density (4) that satisfies a specified error bound ε. In particular, this work seeks to find the structure and parameters $\boldsymbol{\eta}$ of the GM

$$g(x, \boldsymbol{\eta}) = \sum_{i=1}^{L} w_i \mathcal{N}(x; \mu_i, \sigma_i^2) = \sum_{i=1}^{L} g_i(x, \boldsymbol{\eta}_i) \tag{7}$$

where the parameter vector is given by

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1^{\top} & \dots & \boldsymbol{\eta}_i^{\top} & \dots & \boldsymbol{\eta}_L^{\top} \end{bmatrix}^{\top} \tag{8}$$

$$\boldsymbol{\eta}_i = \begin{bmatrix} \beta_i \\ \mu_i \\ \sigma_i \end{bmatrix}, \ i = 1, \dots, L \tag{9}$$

where $\beta_i^2 = w_i$ are used to ensure that the approximation density's weights never become negative [16].

III. Background on Progressive Bayes

The progressive Bayes [12] framework utilizes a homotopy continuation to solve for the unknown parameters of a known (generally non-Gaussian) density function. The homotopy continuation involves approximating a parameterized target function that progressively becomes more similar to the final desired density. The parameterized target function, hereon referred to as the *representative function* [12] $\tilde{f}(x, \gamma)$, must satisfy two properties with respect to the parameter $\gamma \in [0,1]$: when $\gamma = 0$, the representative function $\tilde{f}(x, \gamma)$ is a function that can be approximated perfectly, and when $\gamma = 1$, $\tilde{f}(x, \gamma)$ is exactly the target function $f(x)$. A representative function candidate is [16]

$$\tilde{f}(x, \gamma) = \gamma f(x) + (1 - \gamma) \hat{f}(x) \tag{10}$$

where $f(x)$ is the target density and $\hat{f}(x)$ is a simple function that can be approximated almost exactly. It is straightforward to prove that this representative function respects the properties mentioned above, regardless of $f(x)$ and $\hat{f}(x)$.

Throughout the progression, it is ensured that $g(x, \boldsymbol{\eta})$ minimizes the L_2 error from the representative function

$$G(\gamma, \boldsymbol{\eta}) = \frac{1}{2} \int_{\mathbb{R}} \left(\tilde{f}(x, \gamma) - g(x, \boldsymbol{\eta}) \right)^2 dx \tag{11}$$

by enforcing the necessary condition for optimality. As shown in [12], [16], [17], the outcome is a set of linear ordinary differential equation (ODE) in γ:

$$\boldsymbol{b}(\boldsymbol{\eta}, \gamma) = \mathbf{P}(\boldsymbol{\eta}, \gamma) \frac{\partial \boldsymbol{\eta}}{\partial \gamma} \tag{12}$$

with

$$\boldsymbol{b}(\boldsymbol{\eta}, \gamma) = \int_{\mathbb{R}} \frac{\partial g(x, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\tilde{f}(x, \gamma)}{\partial \gamma} dx \tag{13}$$

$$\begin{aligned} \mathbf{P}(\boldsymbol{\eta}, \gamma) = \int_{\mathbb{R}} & \frac{\partial g(x, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \left(\frac{\partial g(x, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right)^{\top} \\ & + \left(g(x, \boldsymbol{\eta}) - \tilde{f}(x, \gamma) \right) \frac{\partial^2 g(x, \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\top}} dx \end{aligned} \tag{14}$$

Explicit expressions for the partials in (13) and (14) are available for univariate mixtures and select representative functions in [16].

As the ODE integration progresses, the distance between the approximating density and the target distribution may increase if $g(x, \boldsymbol{\eta})$ lacks a sufficient number of components to accurately model $\tilde{f}(x, \gamma)$. A manipulation of (11) yields a more interpretable relative error percentage

$$d(\gamma, \boldsymbol{\eta}) = 100 \sqrt{\frac{\int_{\mathbb{R}} \left(\tilde{f}(x, \gamma) - g(x, \boldsymbol{\eta}) \right)^2 dx}{\int_{\mathbb{R}} \tilde{f}(x, \gamma)^2 dx + \int_{\mathbb{R}} g(x, \boldsymbol{\eta})^2 dx}} \tag{15}$$

When $d(\gamma, \boldsymbol{\eta}) \geq \varepsilon$, the progression is paused, and the GM resolution is increased in key regions in a process known as *structural adaptation* [12], [16]. The component contributing the greatest error is referred to as the *critical* component and is defined by the index

$$c = \arg \max_{i=1,\ldots,L} \{ G_i(\gamma, \boldsymbol{\eta}) \} \tag{16}$$

$$G_i(\gamma, \boldsymbol{\eta}) = \int_{\mathbb{R}} \left(\tilde{f}(x, \gamma) - g(x, \boldsymbol{\eta}) \right)^2 g_i(x, \boldsymbol{\eta}_i) dx \tag{17}$$

The critical component is then replaced by a GM and is essentially split into smaller components, as discussed in Section IV-C. While the increased GM resolution enables a better approximation of the representative function, the split density approximates only the lower-resolution GM and requires small adjustments $\Delta\boldsymbol{\eta}$ to better approximate the representative function.

References [12], [16] propose an iterative *correction step*

$$\boldsymbol{\eta}_{k+1} = \boldsymbol{\eta}_k + \Delta\boldsymbol{\eta}_k, \ k = 0, \ldots, k_{\max} \tag{18}$$

where

$$\left. \frac{\partial^2 G(\gamma, \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\top}} \right|_{\boldsymbol{\eta}_k} \Delta\boldsymbol{\eta}_k = \left. \frac{\partial G(\gamma, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}_k} \tag{19}$$

The error Hessian in (19) is equivalent to the matrix $\mathbf{P}$ defined in (14), and

$$\frac{\partial G(\gamma, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \int_{\mathbb{R}} \left(\tilde{f}(x, \gamma) - g(x, \boldsymbol{\eta}) \right) \frac{\partial g(x, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} dx \tag{20}$$

The iterative correction procedure is terminated and the progression resumes when either $\Delta\boldsymbol{\eta} \to \mathbf{0}$ or when $k = k_{\max}$, whichever occurs first. In some cases, the contribution of a component may become negligible later in the progression. In that case, the insignificant component may be deleted or pruned from the mixture to improve computational efficiency.

IV. Progressive Chernoff Fusion

This section presents a novel method for finding a GM approximation to the Chernoff fusion, or more generally, EMD fusion, posterior with accuracy guarantees. A simple method for computing the optimal Chernoff weight is introduced and shown to be more accurate and computationally efficient on average compared to existing approaches. Then, a progressive Bayes-based solution is proposed, which transforms a first-order approximation of the Chernoff posterior to a high accuracy GM approximation that satisfies a specified error tolerance. This section also discusses structural adaptation and proposes multiple enhancements for improving convergence.

A. Optimal Fusion Parameter Computation: Root Finding Optimization Method

In EMD fusion, the problem of choosing the "best" fusion weight ω is the subject of ongoing research [18]. A reasonable choice is to balance the information contributed by each input distribution, and thus this paper adopts the Chernoff weight as defined in (5). Because no closed-form solution of ω^* is available, numerical solutions are required.

This paper proposes a simple root-finding based solution to the optimal Chernoff fusion weight. The first and second derivatives of $J_\omega(\omega)$ with respect to ω are

$$\frac{\partial J_\omega(\omega)}{\partial \omega} = \int_{\mathbb{R}} \frac{\partial}{\partial \omega} \left(p_1(x)^\omega p_2(x)^{1-\omega} \right) dx \tag{21}$$

$$\frac{d^2 J_\omega(\omega)}{d\omega^2} = \int_{\mathbb{R}} p_1(x)^\omega p_2(x)^{1-\omega} \log \left(\frac{p_1(x)}{p_2(x)} \right)^2 dx \geq 0 \tag{22}$$

The second derivative is clearly non-negative and thus $J_\omega(\omega)$ is convex, meaning any local minimum of J_ω is a global minimum. By the necessary condition of optimality, ω^* satisfies

$$\int_{\mathbb{R}} p_1(x)^{\omega^*} p_2(x)^{1-\omega^*} \log \left(\frac{p_1(x)}{p_2(x)} \right) dx = 0 \tag{23}$$

While a closed-form solution to the integral in (23) is unavailable, the integral can be well approximated with a convergence rate of $\mathcal{O}(N_s^{-1/2})$ [19] by Monte Carlo integration as

$$\begin{aligned} & \int_{\mathbb{R}} p_1(x) \left(\frac{p_2(x)}{p_1(x)} \right)^{1-\omega^*} \log \left(\frac{p_1(x)}{p_2(x)} \right) dx \approx \\ & \sum_{s=1}^{N_s} \left(\frac{p_1(x_s)}{p_2(x_s)} \right)^{\omega^*-1} \log \left(\frac{p_1(x_s)}{p_2(x_s)} \right) = 0 \end{aligned} \tag{24}$$

where $\{x_s\}_{s=1}^{N_s} \sim p_1(x)$. By this approximation, ω^* is the unique root of (24) and is solvable using simple numerical root-finding methods, such as Newton-Raphson and Halley's methods. This paper adopts Halley's method to leverage the readily available first and second-order derivatives of (24) for faster convergence. This method, hereon referred to as the *root finding optimization algorithm (RFOpt)*, is compared to other existing approaches in Section V. Once ω^* is obtained, the exact fusion rule is established. In all but simple cases, however, the weighted exponential product of two GMs does not yield a GM, motivating the approximation introduced in the following subsections.

B. Chernoff Representative Function

The key objective of this work is to obtain an accurate approximation of the Chernoff density. Assuming the optimal fusion parameter ω^* is known (or approximated sufficiently well), a natural choice for the target density is the unnormalized density

$$f(x) = p_1(x)^{\omega^*} p_2(x)^{1-\omega^*} \tag{25}$$

which admits simpler derivatives than its normalized counterpart. As discussed in Section III, the representative function defines the progression from some tractable density to the true target density. One candidate representative function for Chernoff fusion is

$$\tilde{f}_{\text{EMD}}(x,\gamma) = p_1(x)^{\gamma\omega^*} p_2(x)^{1-\gamma\omega^*} \tag{26}$$

By this representative function, the progression begins with $\tilde{f}(x,\gamma = 0)$ equal to the known density $p_2(x)$. Then the solution continuously approaches the Chernoff fusion target via intermediate EMDs. Inspired by (10), another candidate representative function is

$$\tilde{f}_{\text{FOPA}}(x,\gamma) = \gamma f(x) + (1-\gamma) f_{\text{FOPA}}(x) \tag{27}$$

where at $\gamma = 0$, the representative function is equal to the so-called first order power approximation (FOPA) $f_{\text{FOPA}}(x)$ from [10]

$$f_{\text{FOPA}}(x) = \sum_{i}^{L} w_i^{\text{FOPA}}(\omega)\mathcal{N}(x;\mu^{\text{FOPA}}(\omega),(\sigma^{\text{FOPA}}(\omega))^2) \tag{28}$$

$$(\sigma_i^{\text{FOPA}}(\omega))^2 = \left(\omega\sigma_{1,r}^{-2} + (1-\omega)\sigma_{2,l}^{-2}\right)^{-1} \tag{29}$$

$$\mu_i^{\text{FOPA}}(\omega) = \sigma_i(\omega)^2\left(\omega\sigma_{1,r}^{-2}\mu_{1,r} + (1-\omega)\sigma_{2,l}^{-2}\mu_{2,l}\right) \tag{30}$$

$$w_i^{\text{FOPA}}(\omega) = \frac{w_{1,r}^{\omega}\sigma_{1,r}^{1-\omega}w_{2,l}^{1-\omega}\sigma_{2,l}^{\omega}\mathcal{N}\left(\mu_{1,r};\mu_{2,l},\frac{\sigma_{1,r}^2}{\omega}+\frac{\sigma_{2,l}^2}{1-\omega}\right)}{u_i(\omega)} \tag{31}$$

$$u_i(\omega) = \sum_{m=1}^{M}\sum_{n=1}^{N} w_{1,m}^{\omega}\sigma_{1,m}^{1-\omega}w_{2,n}^{1-\omega}\sigma_{2,n}^{\omega} \cdot \mathcal{N}\left(\mu_{1,m};\mu_{2,n},\frac{\sigma_{1,m}^2}{\omega}+\frac{\sigma_{2,n}^2}{1-\omega}\right) \tag{32}$$

$$i = (r-1)N + l, \quad r = 1,\ldots,M, \quad l = 1,\ldots,N \tag{33}$$

FOPA is a straightforward and computationally inexpensive approximation of the Chernoff fusion rule, although its accuracy may be insufficient in some fusion applications, especially when the mixture components are close to each other [10]. While the FOPA-based initialization is generally closer to $f(x)$ compared to $p_2(x)$, the approximation $g(x,\boldsymbol{\eta})$ inherits the larger NM mixture size, which is undesirable for computational efficiency. The third and final representative function proposed in this work is a variant of the FOPA-based progression, where the progression is initialized at the moment-matched Gaussian approximation of the FOPA. This FOPA moment matched (FMM) representative function is defined as

$$\tilde{f}_{\text{FMM}}(x,\gamma) = \gamma f(x) + (1-\gamma) f_{\text{FMM}}(x) \tag{34}$$

where

$$f_{\text{FMM}}(x) = \mathcal{N}(x;\,\mu^{\text{FMM}}(\omega),\,(\sigma^{\text{FMM}}(\omega))^2) \tag{35}$$

$$\mu^{\text{FMM}}(\omega) = \sum_{i=1}^{L} w_i^{\text{FOPA}}(\omega)\mu_i^{\text{FOPA}}(\omega) \tag{36}$$

$$(\sigma^{\text{FMM}}(\omega))^2 = \sum_{i=1}^{L} w_i^{\text{FOPA}}(\omega)\left((\sigma_i^{\text{FOPA}}(\omega))^2 + (\mu_i^{\text{FOPA}}(\omega))^2\right) - (\mu^{\text{FMM}}(\omega))^2 \tag{37}$$

The candidate representative functions (27) and (34) yield partial derivatives with fewer numerical sensitivities compared to (26). For this reason and its advantageous smaller mixture size, (34) is used exclusively in the remainder of this work.

C. Expanded Structural Adaptation

At $\gamma = 0$, $g(x,\boldsymbol{\eta})$ is an exact approximation of the representative function. As γ is varied, the parameters $\boldsymbol{\eta}$ are adapted continuously to maintain an optimal approximation of $\tilde{f}(x,\gamma)$. At certain points in the progression, however, the finite number of components in $\boldsymbol{\eta}$ may be insufficient to maintain the desired error bound, in which case additional components may be introduced. Similarly, some components may become negligible to the approximation and may be pruned from the mixture. This process of component addition and deletion is referred to as *structural adaptation* and was first proposed in [12]. This subsection expands the original structural adaptation algorithm and proposes a new line-search based correction step and dynamic split size correction to improve algorithmic stability.

The critical component, identified according to (16), is split following a splitting library that minimizes the L_2 distance between the univariate standard Gaussian $\pi(x) = \mathcal{N}(x;0,1)$ and $\bar{\pi}(x)$, a distribution of R components. The said distribution is then shifted and scaled to approximate the critical component as

$$g_c(x,\boldsymbol{\eta}_c) \approx \sum_{i=1}^{R} g_{c,i}(x,\bar{\boldsymbol{\eta}}_{c,i}) \tag{38}$$

$$\bar{\boldsymbol{\eta}}_{c,i} = \begin{bmatrix}\beta_{c,i} & \mu_{c,i} & \sigma_{c,i}\end{bmatrix}^{\top} \tag{39}$$

$$w_{c,i} = \bar{w}_i w_c \tag{40}$$

$$\mu_{c,i} = \mu_c + \sigma_c \bar{\mu}_i \tag{41}$$

$$\sigma_{c,i} = \bar{\sigma}\sigma_c \tag{42}$$

where $\bar{w}_i$, $\bar{\mu}_i$, and $\bar{\sigma}$ are the parameters of the univariate standard Gaussian split library. Further discussion on split library generation and applications can be found in [12], [20], [21].

Once a component is split into smaller components, an additional correction step is required to reduce the overall approximation error $G(\gamma, \boldsymbol{\eta})$. The original correction step (18) does not necessarily lead to an improved approximation, in which case the progression can diverge. This stability issue is addressed in the proposed expanded structural adaptation stage by two modifications.

The first improvement is a gradient descent correction

$$\boldsymbol{\eta}_{k+1} = \boldsymbol{\eta}_k + \alpha_k \Delta\boldsymbol{\eta}_k, \;\; k = 0, \ldots, k_{\max} \tag{43}$$

where the learning parameter α_k controls the step size and is reduced if the proposed step leads to increased error. In other words, the gradient descent is obtained by checking that the error at the $(k+1)^{\text{th}}$ step, $G_{k+1} < G_k$; otherwise q is incremented to reduce the step size as

$$\alpha_k = \frac{1}{q}, \;\; q = 1, ..., q_{\max} \tag{44}$$

where $q_{\max}$ is the maximum number of reductions allowed before proceeding to the next step. For simplicity, this paper chooses $q_{\max} = k_{\max} = 10$. The correction stage is terminated when either the Euclidean norm $\|\Delta\boldsymbol{\eta}\|_2 < 10^{-5}$, or $k = k_{\max}$, whichever occurs first.

The second structural adaptation improvement involves a dynamic split size adjustment in the event that the initial adaptation fails to meet the user-imposed error threshold. As described in Algorithm 1, if the original $R = 2$ component split does not result in the tolerance being met, the critical component is replaced by an $R = 3$ component split, and so on.

Algorithm 1: Expanded structural adaptation.

Input: $p_1(x)$; $p_2(x)$; ω^*; γ; $\boldsymbol{\eta}$; ε
Output: $\boldsymbol{\eta}$

```
Identify critical component c from (16)
Compute f̃(x, γ) from (7)
Obtain d(γ, η) from (15)
R ← 2
while d(γ, η) > ε do
    η̄ ← split g_c(x) into R components
    Replace η_c with η̄ in η
    η_0 ← η
    η ← correction_step(η_0)
    Obtain d(γ, η) from (15)
    R ← R + 1
end
```

V. Numerical Results

This section demonstrates the effectiveness of the new progressive Chernoff fusion in a series of numerical examples. First, the RFOpt method for finding the optimal Chernoff fusion weight is compared to three existing state-of-the-art methods. Then, the full progressive Chernoff algorithm is demonstrated in two examples involving non-Gaussian input densities and a variety of desired error tolerances.

A. RFOpt Performance

To evaluate the performance of RFOpt, three other approaches are considered and summarized here. The first alternative is the FMM determinant method (FMM-Det) proposed in [7], which chooses ω^* to be the minimizer of the determinant of the FOPA approximation covariance, as given in (37). The second alternative, importance sampling-based optimization algorithm (ISOpt) [8], uses a golden search over ω, where intermediate objective function evaluations are computed using importance sampling Monte Carlo integration with a first-order approximate proposal density (computed for $\omega = 0.5$). The last alternative considered is a brute-force grid-based search method over ω which, while impractical in most settings, provides an accuracy benchmark for the sake of comparison. In particular, 10^6 evenly spaced points between 0 and 1 are used in the following example.

The two input GMs $p_1(x)$ and $p_2(x)$ [17] are defined in Table I and shown in Fig. 2. Each method's solution $\hat{\omega}$

TABLE I
Example 1 input distributions mixand values.

	$p_1(x)$			$p_2(x)$		
i	w	μ	σ	w	μ	σ
1	0.2	−3	1	1	−2	1
2	0.4	−1	1			
3	0.2	1	1			
4	0.2	3	1			

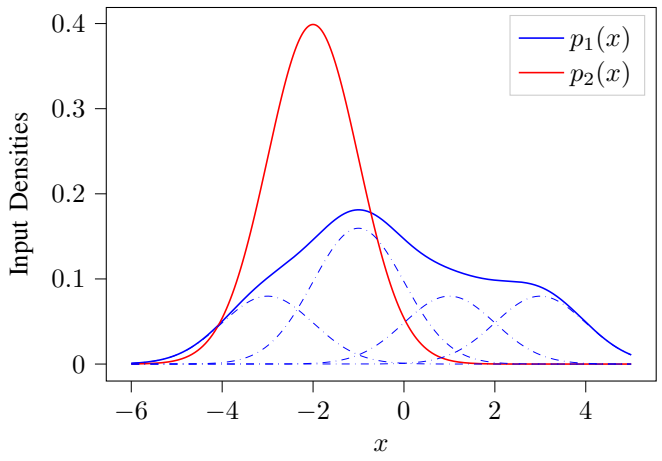

Fig. 2. Example 1 inputs pdfs $p_1(x)$ (solid blue) and $p_2(x)$ (solid red), where the constituent mixands of $p_1(x)$ are represented by the dash-dotted blue lines.

is recorded and its suboptimality is measured by its corresponding cost $J_\omega(\hat{\omega})$ as defined by (6). The percentage error is defined as the normalized relative difference of each algorithm's minimal with respect to grid-based optimal cost $\hat{e} = (J_{\hat{\omega}}(\omega) - \hat{J}_{\text{GB}})/\hat{J}_{\text{GB}}$. Because ISOpt and RFOpt are inherently stochastic, Table II reports the performance metrics averaged over 5,000 independent Monte Carlo trials. The FMM-Det approach was the most computationally efficient method tested but produced significant errors. The RFOpt algorithm achieved the lowest average error, producing lower costs than the ISOpt approach in 2,817 of the 5,000 trials. The RFOpt also achieved the fastest runtimes after the FMM-Det approach, producing solutions at twice the speed of ISOpt on average. While more conclusive findings require expanded testing on a variety of univariate and multivariate distributions, the preliminary results reported here suggest that the simple RFOpt method is fast and highly accurate compared to existing methods.

TABLE II
COMPARISON OF CHERNOFF FUSION WEIGHT OPTIMIZATION APPROACHES, WHERE VALUES REPORTED FOR STOCHASTIC METHODS ARE AVERAGED OVER $5,000$ INDEPENDENT TRIALS.

Algorithm	$J_\omega(\hat{\omega})$	$\hat{e}$ [%]	$\hat{\omega}$	Time [s]
Grid-Based	0.75550	–	0.6925	33.9637
FMM-Det	1.00000	32.3613	0.0	0.0029
ISOpt	0.75590	0.0516	0.6941	0.01472
RFOpt	0.75574	0.0318	0.6923	0.0086

B. Progressive Chernoff Performance

To demonstrate the effectiveness of the progressive Chernoff fusion method, two examples are presented. The first example involves the fusion of a Gaussian and GM density as shown in Fig. 2. A more complicated fusion problem is considered in the second example, where the input densities are both non-Gaussian, as shown in Fig. 6. The Runge–Kutta–Fehlberg method [22] is used in each experiment to perform the ODE integration of (12).

Using an error tolerance of $\varepsilon = 3\%$, the progressive Chernoff fusion solution for the Example 1 pdf is a Gaussian density with $\mu = -1.60$, and $\sigma = 1.42$ and total approximation error of 2.06%. Fig. 3 shows the comparison between the progressive Chernoff approximation and the true target density.

A more challenging case is then considered by repeating the experiment with a strict error threshold of 0.1%. A single Gaussian component approximation is incapable of satisfying the tighter error tolerance, resulting in multiple expanded structural adaptation stages throughout the progression. The resulting seven-component GM approximation, which ultimately achieves an error of 0.0527%, is tabulated in Table III and visually compared to the target density in Fig. 4. The approximation error throughout the progression is shown in Fig. 5 and clearly illustrates how expanded structural adaptation reduces the error in multiple instances.

The last experiment considers the Example 2 input densities given in Table IV, and which are both non-Gaussian. Using

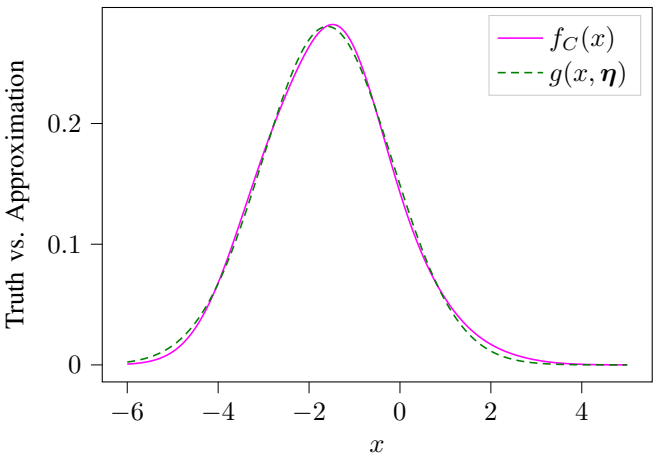

Fig. 3. Progressive Chernoff fusion algorithm with 3% allowed maximum error, resulting in 2.06% error.

TABLE III
GM DISTRIBUTION RESULT WHEN ENFORCING A MAXIMUM ERROR OF 0.1%.

i	w	μ	σ	i	w	μ	σ
1	0.02	−3.57	1.02	5	0.20	−1.19	0.77
2	0.24	−0.49	1.02	6	0.14	−3.23	0.85
3	0.06	0.83	1.20	7	0.15	−2.43	0.87
4	0.18	−1.91	0.89				

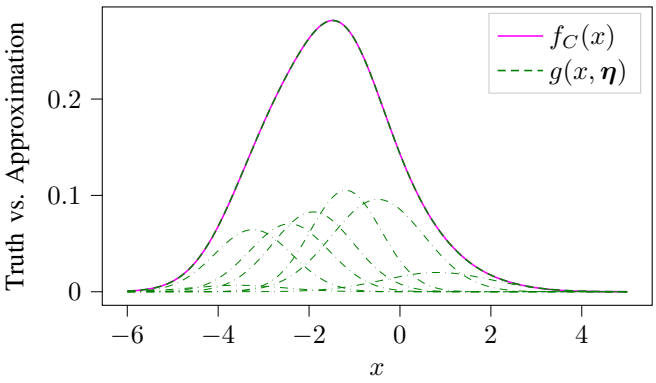

Fig. 4. Progressive Chernoff fusion algorithm with 0.1% allowed maximum error, resulting in 0.0527% error.

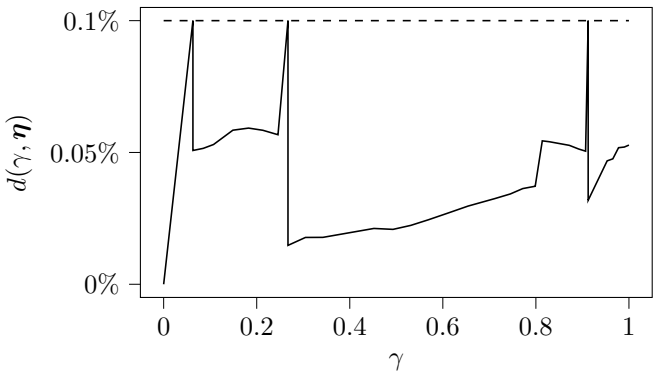

Fig. 5. Error history throughout the homotopy continuation, where the dashed line represents the specified error tolerance.

a maximum error of 5%, the progressive Chernoff algorithm produces a four-component GM, whose values are provided in Table V. As shown in Fig. 8, the proposed approach accurately captures the true multi-modal target density, achieving a final error of 1.07%.

TABLE IV
EXAMPLE 2 INPUT DISTRIBUTIONS MIXAND VALUES.

	$p_1(x)$			$p_2(x)$		
i	w	μ	σ	w	μ	σ
1	0.2	−0.25	1	0.5	−0.5	0.1
2	0.8	0.5	0.3	0.5	0	0.1

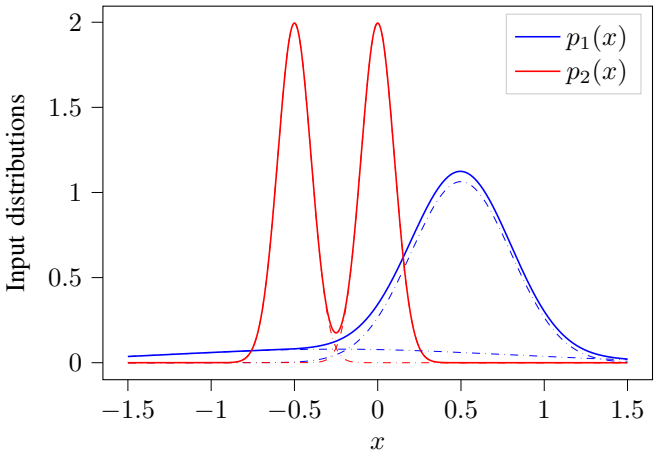

Fig. 6. Example 2 input pdfs $p_1(x)$ (solid blue) and $p_2(x)$ (solid red), where each density's mixands are represented by dash-dotted curves in the corresponding color.

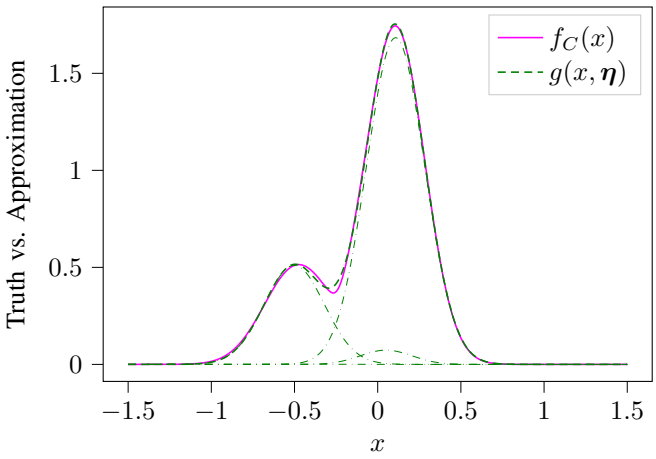

Fig. 7. Progressive Chernoff fusion algorithm with 5% allowed maximum error, resulting in 1.07% error.

VI. CONCLUSIONS AND FUTURE WORK

This paper presents a new method for finding arbitrarily accurate approximations of the Chernoff fusion density based on the Progressive Bayes framework. The performance of the method is demonstrated in multiple example problems

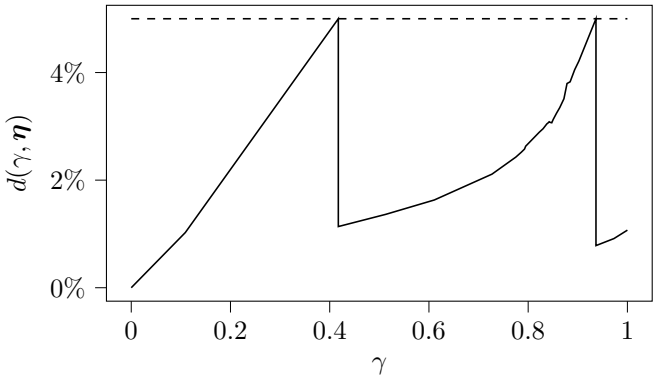

Fig. 8. Error history throughout the homotopy (solid) compared to the maximum tolerated error (dashed).

TABLE V
GM DISTRIBUTION RESULT WHEN ENFORCING A MAXIMUM ERROR OF 5%.

i	w	μ	σ	i	w	μ	σ
1	0.74	0.11	0.18	3	0.23	0.18	0.18
2	0.00	0.18	0.18	4	0.03	0.05	0.15

involving the fusion of non-Gaussian densities and is shown to be highly accurate. Future work will consider extensions of the method to multivariate distributions and investigate alternative integration methods for improved computational efficiency. With further development, the proposed method could be applied to a wide range of distributed estimation problems.

VII. ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their knowledgeable insights, which greatly improved the quality of the manuscript.

REFERENCES

[1] Ondrej Hlinka, Franz Hlawatsch, and Petar M. Djuric. Distributed particle filtering in agent networks: A survey, classification, and comparison. *IEEE Signal Processing Magazine*, 30(1):61–81, January 2013.

[2] Augustus Buonviri, Matthew York, Keith LeGrand, and James Meub. Survey of Challenges in Labeled Random Finite Set Distributed Multi-Sensor Multi-Object Tracking. In *2019 IEEE Aerospace Conference*, pages 1–12, March 2019. ISSN: 1095-323X.

[3] Chee-Yee Chong, Kuo-Chu Chang, and Shozo Mori. Fundamentals of Distributed Estimation. In *Distributed Data Fusion for Network-Centric Operations*. CRC Press, 2013. Num Pages: 30.

[4] Tim Bailey, Simon Julier, and Gabriel Agamennoni. On Conservative Fusion of Information with Unknown Non-Gaussian Dependence. *Proceedings of the 15th International Conference on Information Fusion Fusion 2012*, pages 1–8, 2012.

[5] Shane Lubold and Clark N. Taylor. Formal definitions of conservative probability distribution functions (PDFs). *Information Fusion*, 88:175–183, December 2022.

[6] M.B. Hurley. An information theoretic justification for covariance intersection and its generalization. In *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997)*, volume 1, pages 505–511 vol.1, July 2002.

[7] Simon J. Julier. An Empirical Study into the Use of Chernoff Information for Robust, Distributed Fusion of Gaussian Mixture Models. In *2006 9th International Conference on Information Fusion*, pages 1–8, July 2006.

[8] Nisar R. Ahmed and Mark Campbell. Fast Consistent Chernoff Fusion of Gaussian Mixtures for Ad Hoc Sensor Networks. *IEEE Transactions on Signal Processing*, 60(12):6739–6745, December 2012.

[9] Nisar R. Ahmed. What's one mixture divided by another?: A unified approach to high-fidelity distributed data fusion with mixture models. In *2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 289–296, September 2015.

[10] Melih Gunay, Umut Orguner, and Mubeccel Demirekler. Chernoff fusion of gaussian mixtures based on sigma-point approximation. *IEEE Transactions on Aerospace and Electronic Systems*, 52(6):2732–2746, December 2016. Conference Name: IEEE Transactions on Aerospace and Electronic Systems.

[11] Jiří Ajgl, Miroslav Šimandl, and Jindřich Duník. Approximation of powers of gaussian mixtures. In *2015 18th International Conference on Information Fusion (Fusion)*, pages 878–885, July 2015.

[12] Uwe D. Hanebeck, Kai Briechle, and Andreas Rauh. Progressive Bayes: a new framework for nonlinear state estimation. In *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003*, volume 5099, pages 256–267. SPIE, April 2003.

[13] William J. Farrell and Chidambar Ganesh. Generalized chernoff fusion approximation for practical distributed data fusion. In *2009 12th International Conference on Information Fusion*, pages 555–562, July 2009.

[14] Nisar Ahmed. Conditionally factorized DDF for general distributed bayesian estimation. In *2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*, pages 1–7, September 2014.

[15] Nisar Ahmed, Jonathan Schoenberg, and Mark Campbell. Fast weighted exponential product rules for robust general multi-robot data fusion. In *Robotics: Science and Systems*, volume 8, pages 9–16, 2013.

[16] Marco F. Huber and Uwe D. Hanebeck. Progressive Gaussian mixture reduction. In *2008 11th International Conference on Information Fusion*, pages 1–8, June 2008.

[17] O.C. Schrempf, O. Feiermann, and U.D. Hanebeck. Optimal mixture approximation of the product of mixtures. In *2005 7th International Conference on Information Fusion*, volume 1, pages 8 pp.–, July 2005.

[18] Clark N. Taylor and Adrian N. Bishop. Homogeneous functionals and bayesian data fusion with unknown correlation. *Information Fusion*, 45:179–189, January 2019.

[19] Russel E. Caflisch. Monte carlo and quasi-monte carlo methods. *Acta Numerica*, 7:1–49, 1998.

[20] Kyle J. DeMars, Robert H. Bishop, and Moriba K. Jah. Entropy-Based Approach for Uncertainty Propagation of Nonlinear Dynamical Systems. *Journal of Guidance, Control, and Dynamics*, 36(4):1047–1057, July 2013.

[21] Keith A. LeGrand and Silvia Ferrari. Split Happens! Imprecise and Negative Information in Gaussian Mixture Random Finite Set Filtering. *Journal of Advances in Information Fusion*, 17(2):78–96, December 2022.

[22] E. Fehlberg. Classical Fifth-, Sixth-, Seventh-, and Eighth-Order Runge-Kutta Formulas with Stepsize Control. Technical Report NASA-TR-R-287, October 1968. NTRS Author Affiliations: NASA Marshall Space Flight Center NTRS Document ID: 19680027281 NTRS Research Center: Legacy CDMS (CDMS).