On Conservative Fusion of Information with Unknown Non-Gaussian Dependence

Tim Bailey
University of Sydney
Email: tbailey@acfr.usyd.edu.au

Simon Julier
University College London
Email: s.julier@cs.ucl.ac.uk

Gabriel Agamennoni
University of Sydney
Email: g.agamennoni@acfr.usyd.edu.au

***Abstract*—This paper examines the notions of consistency and conservativeness for data fusion involving dependent information, where the degree of dependency is unknown. We consider these notions in a general sense, for non-Gaussian probability distributions, in terms of structural consistency and information processing, in particular the counting of common information. We consider the role of entropy in defining a conservative fusion rule. Finally, we investigate the *geometric mean density* (GMD) as a particular fusion rule, which generalises the *Covariance Intersection* rule to non-Gaussian pdfs. We derive key properties to demonstrate that the GMD is both conservative and effective in combining information from dependent sources.**

***Keywords*—Conservative fusion, non-Gaussian, geometric mean density, double counting**

I. Introduction

Data fusion within the Bayesian framework requires dependence between uncertain variables to be defined in terms of joint probability density functions (pdfs) or conditional pdfs. However, for some estimation problems, such as decentralised data fusion (e.g., [18], [1]), maintaining the joint structure incurs significant bookkeeping and communications overhead and imposes strong constraints on network topology. Feasible implementation may necessitate suboptimal fusion procedures that discard some dependency information. Similar problems arise when combining information from multiple sources that include unknown amounts of correlation, such as systematic modelling errors. These fusion methods necessarily deviate from the standard Bayesian form.

This paper considers the problem of data fusion when the dependence between pdfs is unknown or has been discarded. We wish to properly account for this dependency, and to define a notion of "*conservativeness*" so as to avoid becoming overconfident; to be able to reuse old information without doing violence to the *shape* of our uncertainty over the parameter space. While the idea of a conservative upper bound on uncertainty (i.e., covariance) has been studied previously in terms of Gaussian pdfs [13], [7], the structure of the Gaussian function—unimodal and symmetric, with coincident mean, median and mode—constrains its behaviour and thereby limits the generality of such a bound. In this paper, we consider non-Gaussian distributions and present phenomena that do not appear with Gaussians. Perhaps most surprising of these is to note that, for optimal Bayesian data fusion, even with independent data, entropy can sometimes increase. With these insights, we make progress towards a general definition of *conservativeness* in terms of point-wise lower bounds, the double counting of common information, entropy increase and shape preservation.

Our investigation focuses on a particular form of conservative fusion rule: the *geometric mean density* (GMD), which is a generalisation of the Gaussian fusion rule known as *Covariance Intersection* (CI) [13]. The two-component form of the GMD rule, which fuses two pdfs $p_a(\mathbf{x})$ and $p_b(\mathbf{x})$, is as follows

$$p_c(\mathbf{x}) = \frac{1}{\eta_c} p_a(\mathbf{x})^{\omega} p_b(\mathbf{x})^{1-\omega} \tag{1}$$

where $0 \leq \omega \leq 1$ and

$$\eta_c = \int p_a(\mathbf{x})^{\omega} p_b(\mathbf{x})^{1-\omega} d\mathbf{x} \tag{2}$$

is the normalisation constant. In general, the GMD may be defined in terms of M components

$$p_c(\mathbf{x}) = \frac{1}{\eta_c} \prod_i^M p_i(\mathbf{x})^{\omega_i} \tag{3}$$

where $\omega_i \geq 0$ and $\sum_i^M \omega_i = 1$. However, for simplicity, we only consider the two component case for the rest of this paper.

The format of this paper is as follows. The next section discusses previous work pertaining to consistency, conservativeness and the GMD fusion rule. Section III considers the role of entropy in data fusion and its relationship to conservative fusion. Section IV presents key properties of the GMD which indicate that it is a conservative fusion rule, and also that it is a *good* fusion rule insofar as it permits information gain. These properties offer insights into sufficient conditions for consistent and conservative fusion of data with unknown dependence. Section V concludes with a summary of the key results.

II. Related Work

In this section we examine previous work on the notion of Bayesian consistency, on conservative fusion in the context of Gaussian distributions and CI, and various applications of the GMD rule related to data fusion.

A. Bayesian Consistency

The Bayesian sum and product rules, and inference according to Bayes' theorem, is justified by Jaynes [8, Sections 1.7, 2.1] as having "*structural consistency*" insofar as they generate *unique* solutions. Given the same data, models and assumptions, the same results are obtained regardless of how the information is combined (e.g., the results are independent of fusion order). Zellner [19] considers Bayesian methods from the perspective of information processing, and notes that, according to a particular set of information measures, Bayes' theorem is "100% efficient"; it is a processing rule that entirely preserves information content.

For Gaussian pdfs, consistency may be expressed in terms of expectations. That is, the estimated covariance matrix $\mathbf{P}$ is consistent if it is equal to the expected second moment of the residual between the estimated mean $\hat{\mathbf{x}}$ and the (unknown) true state $\mathbf{x}_t$,

$$\mathbf{P} = E\left[(\hat{\mathbf{x}} - \mathbf{x}_t)(\hat{\mathbf{x}} - \mathbf{x}_t)^T\right]. \tag{4}$$

However, for non-Gaussian pdfs (e.g., with heavy tails, asymmetry or multiple modes), the second moment may be a poor representative of the true shape of the uncertainty over the parameter space, and so a covariance estimate is not a sufficient measure of consistency in general.

It is possible also to specify general properties of *inconsistency*. A pdf is inconsistent if

$$p(\mathbf{x}_t) = 0, \tag{5}$$

since its domain must include the true state. A rule for *fusing* two pdfs is inconsistent if it counts the same information multiple times, since repeated use of the same information modifies the posterior pdf, violating the uniqueness property of structural consistency.[1] Within the Bayesian framework, it is possible to fuse dependent pdfs so long as one explicitly accounts for the common information. Thus, given two pdfs $p(\mathbf{x}|\mathbf{Z}_a)$ and $p(\mathbf{x}|\mathbf{Z}_b)$ that are dependent due to shared data $\mathbf{Z}_a \cap \mathbf{Z}_b \neq \emptyset$, the normalised product $\frac{1}{\eta} p(\mathbf{x}|\mathbf{Z}_a)\, p(\mathbf{x}|\mathbf{Z}_b)$ is an inconsistent fusion rule because it counts the common information twice. However, provided the common information $p(\mathbf{x}|\mathbf{Z}_a \cap \mathbf{Z}_b)$ is known, it is possible to remove the repeated information to obtain an optimal Bayesian estimate.

Theorem 1: The product of dependent pdfs with division by the common information results in a Bayesian posterior pdf [2]. If $\mathbf{Z}_a = \{\mathbf{z}_1, \mathbf{z}_2\}$ and $\mathbf{Z}_b = \{\mathbf{z}_2, \mathbf{z}_3\}$ denote two sets of data, where $\mathbf{z}_2 = \mathbf{Z}_a \cap \mathbf{Z}_b$ is the common information and the data are conditionally independent such that $p(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3|\mathbf{x}) = \prod_i p(\mathbf{z}_i|\mathbf{x})$, then

$$p(\mathbf{x}|\mathbf{Z}_a \cup \mathbf{Z}_b) \propto \frac{p(\mathbf{x}|\mathbf{Z}_a)\, p(\mathbf{x}|\mathbf{Z}_b)}{p(\mathbf{x}|\mathbf{Z}_a \cap \mathbf{Z}_b)} \tag{6}$$

Proof: Applying Bayes' theorem to each part of the right-hand-side,

$$\begin{aligned}
&\frac{p(\mathbf{x}|\mathbf{Z}_a)\, p(\mathbf{x}|\mathbf{Z}_b)}{p(\mathbf{x}|\mathbf{Z}_a \cap \mathbf{Z}_b)} \\
&= \frac{p(\mathbf{z}_1, \mathbf{z}_2|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{z}_1, \mathbf{z}_2)} \times \frac{p(\mathbf{z}_2, \mathbf{z}_3|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{z}_2, \mathbf{z}_3)} \times \frac{p(\mathbf{z}_2)}{p(\mathbf{z}_2|\mathbf{x})\, p(\mathbf{x})} \\
&= \underbrace{\frac{p(\mathbf{z}_2)}{p(\mathbf{z}_1, \mathbf{z}_2)\, p(\mathbf{z}_2, \mathbf{z}_3)}}_{=K} \times \frac{p(\mathbf{z}_1, \mathbf{z}_2|\mathbf{x})\, p(\mathbf{z}_2, \mathbf{z}_3|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{z}_2|\mathbf{x})} \\
&= K \cdot \frac{p(\mathbf{z}_1|\mathbf{x})\, p(\mathbf{z}_2|\mathbf{x})\, p(\mathbf{z}_2|\mathbf{x})\, p(\mathbf{z}_3|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{z}_2|\mathbf{x})} \\
&= K \cdot p(\mathbf{z}_1|\mathbf{x})\, p(\mathbf{z}_2|\mathbf{x})\, p(\mathbf{z}_3|\mathbf{x})\, p(\mathbf{x}) \\
&\propto p(\mathbf{x}|\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)
\end{aligned}$$

where K is a scalar constant given the data (i.e., K is not a function of $\mathbf{x}$). ■

Therefore, repeated fusion of dependent pdfs will give a unique solution equal to the information extracted from a single counting of the data.

B. Conservative Gaussian Fusion

In the context of Gaussian pdfs,[2] conservative data fusion has received attention in the development of the CI algorithm [13]. To fuse two sets of marginal statistics, $\{\hat{\mathbf{x}}_a, \mathbf{P}_a\}$ and $\{\hat{\mathbf{x}}_b, \mathbf{P}_b\}$, over the space of $\mathbf{x}$ when they possess unknown correlations, one applies the following rule,

$$\mathbf{P}_c^{-1} = \omega \mathbf{P}_a^{-1} + (1-\omega)\mathbf{P}_b^{-1}, \tag{7}$$

$$\mathbf{P}_c^{-1}\hat{\mathbf{x}}_c = \omega \mathbf{P}_a^{-1}\hat{\mathbf{x}}_a + (1-\omega)\mathbf{P}_b^{-1}\hat{\mathbf{x}}_b, \tag{8}$$

where $0 \leq \omega \leq 1$. The CI fusion rule is equivalent to first inflating the component covariances as $\frac{\mathbf{P}_a}{\omega}$ and $\frac{\mathbf{P}_b}{1-\omega}$, and then applying the standard Bayesian fusion rule, which is simply a Kalman update operation.

The CI algorithm satisfies a necessary condition of a conservative Gaussian upper-bound

$$\mathbf{P}_c - \mathbf{P} \succcurlyeq 0, \tag{9}$$

where $\mathbf{P}$ is the covariance estimate that would be generated by a Kalman update if the correlations were actually known. A stronger property is that $\mathbf{P}_c$ is an upper-bound on the second moment of the CI estimate errors

$$\mathbf{P}_c \succcurlyeq E\left[(\hat{\mathbf{x}}_c - \mathbf{x}_t)(\hat{\mathbf{x}}_c - \mathbf{x}_t)^T\right]. \tag{10}$$

This property also holds for non-Gaussian pdfs, although as stated previously, it is of limited value if the pdf deviates significantly from a Gaussian shape.

C. Data Fusion and the GMD

The GMD generalises the CI fusion rule to non-Gaussian probability distributions. Mahler [15] first noted that for a Gaussian pdf, $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P})$, the covariance inflation of CI was equivalent to raising the Gaussian function to a power and normalising,

$$\frac{1}{\eta} p(\mathbf{x})^{\omega} = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \frac{\mathbf{P}}{\omega}), \tag{11}$$

[1] Double counting of information causes fixation on a subset of the data, undermining the contribution of other data and corrupting the shape of the posterior pdf.

[2] Both the Kalman filter and CI may also be applied to the moment statistics of non-Gaussian pdfs, including higher-order moments. However, since a non-Gaussian pdf and the pdf represented by its first two moments may differ by an arbitrary amount, we feel justified in referring to the two-moment versions of the Kalman filter and CI (and, indeed, *any* two-moment rule) as *Gaussian* fusion rules. They are exactly representative only for Gaussians.

and so showed that the CI algorithm is a special case of the GMD fusion rule (1).

The GMD appears in the probabilistic literature under various guises, with different names corresponding to different application areas. In the context of decision theory [3, page 312], the GMD is used to compute the *Chernoff bound*, where $p_c(\mathbf{x})$ defines a decision boundary between two hypotheses and ω is chosen to minimise the probability of error by matching the Kullback-Leibler divergence of $p_c(\mathbf{x})$ from the hypothesis priors, so that $D(p_c(\mathbf{x})\,||p_a(\mathbf{x})) = D(p_c(\mathbf{x})\,||p_b(\mathbf{x}))$. For this value of ω, the negative log of the normalisation term $-\ln\eta_c$ is called the *Chernoff information*. More recently, the Chernoff information has been applied to data fusion to select the value of ω for the CI algorithm [7], instead of the traditional utilities of minimising the determinant or trace. The term "Chernoff information" also appears as a synonym for the GMD [10]. Another synonym is *exponential mixture densities* [4], [12], even though the pdfs are combined as a product not a sum. We prefer the name GMD because fusion is accomplished as the weighted geometric mean of the component density functions.

Perhaps the earliest and most prevalent instance of the GMD rule is the *logarithmic opinion pool* [5], which is applied to combining pdfs from multiple sources as a "mixture of experts". This application is essentially the same as ours, with a different emphasis: we emphasise common information, typically in the form of a common prior, while they are concerned with different priors but a shared likelihood function. Logarithmic opinion pools typically optimise ω so as to minimise the Kullback-Leibler divergence between the resultant fused pdf and the component pdfs [6]. This is justified as attempting to minimise the overall deviation from each expert.

A related fusion rule is proposed by O'Brien [16], which has the form

$$p_q(\mathbf{x}) = \frac{1}{\eta_q} p_a(\mathbf{x})^{\alpha}\, p_b(\mathbf{x})^{\beta}, \tag{12}$$

where $\alpha, \beta \in [0,1]$. The difference here is that β is not constrained to equal $1-\alpha$. This rule is described by Zellner [20] as "quality adjusted" inputs, although with only informal justification. We shall see in (18) below that the "summation to one" constraint avoids double counting of common information. This property is not satisfied if α and β can change freely. In particular, $\alpha = \beta = 1$ gives the standard Bayes update, which is valid only if the component pdfs are conditionally independent. Similarly, a choice of weights such that $\alpha+\beta > 1$ implicitly assumes that some degree of dependence is known and can be accounted for [11].

III. Entropy and Conservativeness

While the positive semi-definite property (10) of the CI algorithm applies to non-Gaussian data fusion, we do not consider a covariance bound sufficient justification for a general conservative fusion rule. For example, any pdf can be replaced by another non-Gaussian pdf that has the same mean and greater covariance, but zero probability at the location of the true state $\mathbf{x}_t$. Approximating the original pdf in this way is not consistent in the sense defined in (5). The question then is how to say one pdf is "more conservative" than another. One possibility is to chose the one that is more "*uncertain*" than the other, where a useful measure for uncertainty is the *entropy* [14]. It would appear that entropy and conservativeness are related since higher entropy means that a pdf is flatter, more spread out, diffuse. Over a discrete or bounded domain, the maximum entropy pdf is the uniform distribution. However, as with the covariance upper-bound, having greater entropy is insufficient to guarantee (5) is avoided.

Since entropy alone is not enough, we propose the following as sufficient[3] conditions for defining a conservative approximation of a pdf.

A pdf $\tilde{p}(\mathbf{x})$ is a conservative approximation of another pdf $p(\mathbf{x})$ if it satisfies two properties:

1) **The non-decreasing entropy property,**

$$H(p(\mathbf{x})) \le H(\tilde{p}(\mathbf{x})). \tag{13}$$

2) **The order preservation property that $\forall \mathbf{x}_i, \mathbf{x}_j$,**

$$\tilde{p}(\mathbf{x}_i) \le \tilde{p}(\mathbf{x}_j) \iff p(\mathbf{x}_i) \le p(\mathbf{x}_j). \tag{14}$$

The first property ensures that the uncertainty associated with the distribution cannot decrease. The second property means that the two pdfs have the same essential shape, the same locations of minima and maxima. The relative flatness of these pdfs is, therefore, directly comparable. An important consequence of property two is that the inconsistency of (5) is avoided: $p(\mathbf{x}) > 0 \implies \tilde{p}(\mathbf{x}) > 0$.

Theorem 2: Let $p(\mathbf{x})$ be a continuous pdf over an unbounded domain[4] and let $\tilde{p}(\mathbf{x})$ be a conservative approximation which obeys (14). The approximation is guaranteed to be consistent in the sense that

$$\tilde{p}(\mathbf{x}) = 0 \iff p(\mathbf{x}) = 0. \tag{15}$$

Proof: Suppose there exists an $\mathbf{x}_i$ such that $\tilde{p}(\mathbf{x}_i) = 0$ and $p(\mathbf{x}_i) = \epsilon > 0$. Thus, $\tilde{p}(\mathbf{x}_i) \le \tilde{p}(\mathbf{x}_j)$ for all $\mathbf{x}_j$. However, there exists an $\mathbf{x}_j$ such that $p(\mathbf{x}_j) < \epsilon$, as $p(\mathbf{x})$ goes to zero almost everywhere, (since its integral is finite). Therefore,

$$\exists \mathbf{x}_j : \tilde{p}(\mathbf{x}_i) \le \tilde{p}(\mathbf{x}_j) \text{ and } p(\mathbf{x}_i) > p(\mathbf{x}_j),$$

which is a contradiction of (14). And so, to satisfy the order preservation property, $\tilde{p}(\mathbf{x}) = 0$ if and only if $p(\mathbf{x}) = 0$. ∎

When considering the role of entropy with regard to data fusion, it is worth noting an important distinction between Gaussian and non-Gaussian pdfs. Optimal fusion, via Bayes' theorem, of two Gaussian pdfs always results in a decrease

[3] We do not presently know whether these properties are *necessary* conditions; there may exist alternative sufficient conditions.

[4] This theorem also holds for bounded domains provided $p(\mathbf{x}) < \epsilon$ at some point in the domain (e.g., goes to zero at a boundary).

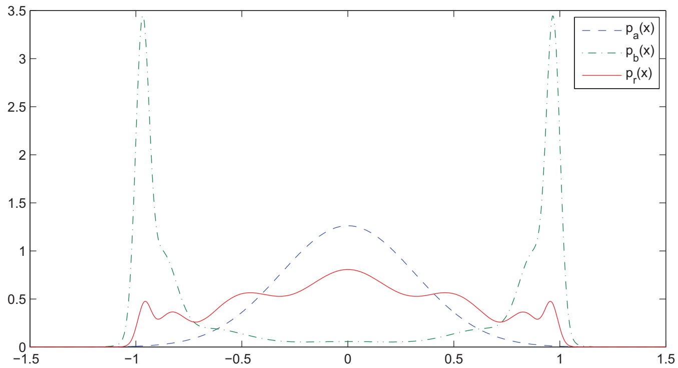

Fig. 1. A Gaussian mixture example showing that Bayesian fusion of two independent pdfs with entropies $H_a = 0.27$ and $H_b = -0.09$, respectively, can generate a pdf with higher entropy, $H_r = 0.66$.

in entropy,[5] which may be interpreted as gaining information or becoming less uncertain. But *optimal* fusion of two non-Gaussian pdfs, may generate a result that has higher entropy than either component. For instance, in Figure 1 we have two Gaussian mixture pdfs where $p_b(\mathbf{x}) \approx 1/p_a(\mathbf{x})$ over the region of significant probability mass. Such pdfs may arise from legitimate sources, be independent and consistent, such that Bayesian fusion is correct and optimal, but the result $p_r(\mathbf{x}) = \frac{1}{\eta} p_a(\mathbf{x}) p_b(\mathbf{x})$ is *more* uncertain because the evidence is conflicting. Fundamentally, Bayesian data fusion is about *combining all available information*, not reducing uncertainty.

IV. Properties of the GMD Fusion Rule

The GMD is compelling as a practical fusion rule and for the insights it provides on conservative fusion in general. In this section, we show that it has properties that appear to satisfy necessary conditions of conservativeness: it does not double count common information and it replaces independent information by conservative approximations. Furthermore, it permits an increase in information, thus making it a *useful* fusion rule when combining information from multiple dependent sources.

A. The GMD is a Conservative Fusion Rule

To show that the GMD is conservative, we first demonstrate that we can make a conservative approximation of pdf $p(\mathbf{x})$ by raising it to a power $p(\mathbf{x})^\omega$, where $0 < \omega \leq 1$. It is trivial from inspection that $p(\mathbf{x})^\omega$ is order preserving

$$p(\mathbf{x}_i) < p(\mathbf{x}_j) \iff p(\mathbf{x}_i)^\omega < p(\mathbf{x}_j)^\omega, \quad \forall \mathbf{x}_i, \mathbf{x}_j. \tag{16}$$

[5]In this instance we are referring to the *differential entropy*, which ranges between $\pm\infty$ (e.g., $H = -\infty$ for a Gaussian with variance $\sigma^2 \to 0$, and $H = \infty$ for a Gaussian with $\sigma^2 \to \infty$). For this paper, having negative entropies is of no special consequence, since we are only interested in ranking uncertainties over a common domain on the continuum from $-\infty$ to $+\infty$, so as to quantify whether one pdf is "more uncertain" than another. The findings of this paper, including the possibility of entropy increasing after optimal fusion, apply equally to discrete probabilities and discrete entropy (which is always positive). Hence, we refer to both differential and discrete entropy simply as "entropy".

Also $p(\mathbf{x})^\omega$ tends towards one at all points thereby increasing the flatness of the function, $|1 - p(\mathbf{x})^\omega| \leq |1 - p(\mathbf{x})|$ for all $\mathbf{x}$. Thus, $p_a(\mathbf{x})^\omega$ has less influence than $p_a(\mathbf{x})$; it exerts less change on $p_b(\mathbf{x})$, so that $p_b(\mathbf{x})$ has shape more similar $p_a(\mathbf{x})^\omega p_b(\mathbf{x})$ than to $p_a(\mathbf{x}) p_b(\mathbf{x})$. The GMD applies this *change reduction* property bilaterally, $p_b(\mathbf{x})^\omega$ also has reduced influence on $p_a(\mathbf{x})$. Furthermore, the normalised pdf $\frac{1}{\eta_\omega} p(\mathbf{x})^\omega$ exhibits an increase in entropy as follows.

Lemma 1: Raising a pdf to a power less than one, and normalising, increases its entropy.

$$H(p_\omega(\mathbf{x})) \geq H(p(\mathbf{x})), \qquad 0 < \omega \leq 1 \tag{17}$$

where $p_\omega(\mathbf{x}) = \frac{1}{\eta_\omega} p(\mathbf{x})^\omega$ and $\eta_\omega = \int p(\mathbf{x})^\omega d\mathbf{x}$.

Proof: The derivative $\frac{dH(p_\omega(\mathbf{x}))}{d\omega}$ is non-positive for all $\omega > 0$ (see the Appendix for details). Therefore the entropy is non-increasing, and has a lower value for $\omega = 1$ than for $0 < \omega < 1$. ■

Note that showing $p(\mathbf{x})^\omega$ is a conservative approximation of $p(\mathbf{x})$ is not sufficient to prove that (1) is conservative. Although each component is replaced by a conservative approximation, the approximations remain dependent on each other, and their normalised product may still be optimistic. However, we can further decompose the GMD into dependent and independent parts, and show that the common information is counted only once.

Theorem 3: The GMD does not double count common information [9].

$$\begin{aligned} p_c(\mathbf{x}) &= \frac{1}{\eta} p(\mathbf{x}|\mathbf{Z}_a)^\omega p(\mathbf{x}|\mathbf{Z}_b)^{1-\omega} \\ &\propto p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x})^\omega p(\mathbf{Z}_b \backslash \mathbf{Z}_a|\mathbf{x})^{1-\omega} p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}) \end{aligned} \tag{18}$$

Proof: The data is presumed to be conditionally independent, such that $p(\mathbf{Z}_a|\mathbf{x}) = p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x}) p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x})$, and so

$$\begin{aligned} & p(\mathbf{x}|\mathbf{Z}_a)^\omega p(\mathbf{x}|\mathbf{Z}_b)^{1-\omega} \\ & \quad \propto (p(\mathbf{Z}_a|\mathbf{x}) p(\mathbf{x}))^\omega (p(\mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}))^{1-\omega} \\ & \quad = p(\mathbf{Z}_a|\mathbf{x})^\omega p(\mathbf{Z}_b|\mathbf{x})^{1-\omega} p(\mathbf{x}) \\ & \quad = (p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x}) p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}))^\omega \\ & \qquad\qquad (p(\mathbf{Z}_b \backslash \mathbf{Z}_a|\mathbf{x}) p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}))^{1-\omega} p(\mathbf{x}) \\ & \quad = p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x})^\omega p(\mathbf{Z}_b \backslash \mathbf{Z}_a|\mathbf{x})^{1-\omega} p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}) \end{aligned}$$

■

Therefore, the GMD has two properties that seem essential to the notion of a general conservative fusion rule.

A fusion rule is conservative if and only if it satisfies two properties:

1) **It does not double count common information,**
2) **It replaces each component of independent information with a conservative approximation.**

A further compelling property of the GMD as a conservative fusion rule is its ability to bound the fused pdf from below.

This property arises from the fact that the normalising constant in (1) is never greater than unity.

Lemma 2: The normalising constant from GMD fusion, $\eta_c(\omega) = \int p_a(\mathbf{x})^{\omega} p_b(\mathbf{x})^{1-\omega} d\mathbf{x}$, is a convex function in ω [12].

Proof: We presume that $p_a(\mathbf{x})$ and $p_b(\mathbf{x})$ are different but have common support. Let $f(\mathbf{x},\omega) = p_a(\mathbf{x})^{\omega} p_b(\mathbf{x})^{1-\omega}$, then the first derivative is

$$\begin{aligned}\frac{\partial f(\mathbf{x},\omega)}{\partial \omega} &= \frac{\partial}{\partial \omega}\left(\frac{p_a(\mathbf{x})}{p_b(\mathbf{x})}\right)^{\omega} p_b(\mathbf{x}) \\ &= f(\mathbf{x},\omega)\ln\frac{p_a(\mathbf{x})}{p_b(\mathbf{x})}, \end{aligned} \quad (19)$$

and the second derivative is

$$\frac{\partial^2 f(\mathbf{x},\omega)}{\partial \omega^2} = f(\mathbf{x},\omega)\left(\ln\frac{p_a(\mathbf{x})}{p_b(\mathbf{x})}\right)^2. \quad (20)$$

This is strictly positive as a function of ω, therefore $f(\mathbf{x},\omega)$ is convex in ω and hence η_c is a convex function of ω. ∎

Since $\eta_c = 1$ at the boundaries, when ω is zero or one, the above lemma implies that $\eta_c \leq 1$ for $0 < \omega < 1$. In addition, a basic property of the weighted geometric mean is that if $x \leq y$ then $x \leq x^{\omega} y^{1-\omega} \leq y$ for any $0 \leq \omega \leq 1$. From these two properties we can show that the GMD is bounded below point-wise by the minimum value of its component pdfs.

Theorem 4: The GMD is bounded below by the minimum of its component functions for all $\mathbf{x}$ [17], [12].

$$p_c(\mathbf{x}) \geq \min\{p_a(\mathbf{x}), p_b(\mathbf{x})\}, \quad \forall \mathbf{x}. \quad (21)$$

Proof: The (non-normalised) weighted geometric mean always lies between the two component functions

$$\begin{aligned}\min\{p_a(\mathbf{x}), p_b(\mathbf{x})\} &\leq p_a(\mathbf{x})^{\omega} p_b(\mathbf{x})^{1-\omega} \\ &\leq \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}, \quad \forall \mathbf{x}.\end{aligned}$$

Therefore, the GMD can fall below $\min\{p_a(\mathbf{x}), p_b(\mathbf{x})\}$ only if $\eta_c > 1$, which is impossible according to Lemma 2. ∎

Remark 1: A consequence of this "bounding below" property is that, if we have multiple dependent pdfs, and each $p_i(\mathbf{x}_0)$ is non-zero at a point $\mathbf{x}_0$, then repeated fusion of dependent information with the GMD cannot cause $p_c(\mathbf{x}_0)$ to diminish to zero at $\mathbf{x}_0$. The component pdf with minimum value at $\mathbf{x}_0$ defines the lowest possible value of $p_c(\mathbf{x}_0)$, so that

$$p_c(\mathbf{x}) \geq \min\{p_i(\mathbf{x})\}_i, \quad \forall \mathbf{x}. \quad (22)$$

It prevents overconfidence in saying "the true state $\mathbf{x}_t$ is *not here*" for some location $\mathbf{x}_0$. This indicates that GMD can be applied *recursively* without tending to generate results that are optimistic in a point-wise sense.

B. The GMD is an Effective Fusion Rule

A conservative fusion rule is not *effective* or *useful* unless it can also improve our knowledge of the uncertain states; it must be able to properly combine the available information. But it is important to realise that to "combine information" does not necessarily mean to "optimise an information measure," such as to reduce entropy. We saw in Section III that there exist cases where optimal Bayesian fusion increases entropy, where combining information increases our measure of uncertainty.

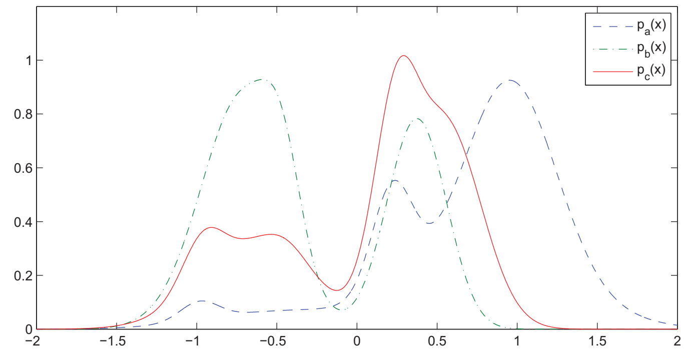

Fig. 2. An example where the GMD rule ($\omega = 0.57$) generates a pdf with a higher maximum peak probability than either component pdf.

Nevertheless, for Bayes' theorem to be a useful fusion rule it must have the *potential* to reduce entropy, even if it does not do so in every case. In this section we show that the GMD is similarly effective insofar as it has the *potential* to optimise an information measure, such as entropy reduction. We also postulate that the GMD behaves sensibly in cases where optimal Bayesian fusion causes entropy to increase.

To show that information gain is *possible* with the GMD it is sufficient to disprove that GMD fails to increase information for *all* instances of $p_a(\mathbf{x})$ and $p_b(\mathbf{x})$. Thus, we have only to provide a single case, a pair of pdfs and a value for ω, for which it increases the information measure. (We do not prove that the GMD *always* increases information, since it clearly does not. One can easily find pdf pairs for which the GMD gives no increase in these measures, regardless of ω. Of course, the GMD can always *equal* that of the best component pdf by simply choosing $\omega = 0$ or $\omega = 1$.)

The following theorems demonstrate that the GMD has the potential to increase information content according to two measures: the maximum peak probability[6] and the entropy.

Theorem 5: The GMD has the potential to increase the maximum peak probability. There exist pairs of pdfs $p_a(\mathbf{x})$ and $p_b(\mathbf{x})$ such that, for some choice of ω,

$$\max p_c(\mathbf{x}) > \max\{\max p_a(\mathbf{x}), \max p_b(\mathbf{x})\} \quad (23)$$

Proof: It is trivial to generate a single empirical example, as shown in Figure 2, thereby proving existence. However, (23) is not true for 1-D Gaussian pdfs, which will always be optimal when $\omega = 0$ or $\omega = 1$. ∎

Theorem 6: The GMD has the potential to decrease entropy. There exist pairs of pdfs $p_a(\mathbf{x})$ and $p_b(\mathbf{x})$ such that, for some choice of ω,

$$H(p_c(\mathbf{x})) < \min\{H(p_a(\mathbf{x})), H(p_b(\mathbf{x}))\} \quad (24)$$

[6] Choosing ω to maximise the peak probability of the GMD is suggested by Mahler [15]. It is worth noting that the choice of ω that maximises peak probability is not the same, in general, as the ω that minimises entropy, but these properties do coincide for Gaussian pdfs.

Proof: This property is also trivial from empirical example, and is well-known for the special case of CI. It is not possible to decrease entropy for 1-D Gaussians; the optimum will always be $\omega = 0$ or $\omega = 1$. ■

Further to these results, our empirical experiments[7] indicate that the potential for the GMD to decrease entropy exists only if optimal Bayesian fusion of the same pdfs results in an entropy decrease. We therefore hypothesise that if optimal Bayesian fusion, with dependencies known and accounted for, causes entropy to increase, then the GMD of the component pdfs, where dependencies are neglected, is unable to decrease entropy, and so the minimum entropy GMD occurs at $\omega = 0$ or $\omega = 1$.

Conjecture 1: If optimal Bayesian fusion does not decrease entropy, then the lower bound entropy for the GMD is that of a component pdf. Suppose we have three independent pdfs $p_i(\mathbf{x})$, $p_j(\mathbf{x})$ and $p_k(\mathbf{x})$. Optimal fusion of these components is given by

$$p_r(\mathbf{x}) = \frac{1}{\eta_r} p_i(\mathbf{x}) p_j(\mathbf{x}) p_k(\mathbf{x}). \tag{25}$$

Suppose further that we have combined estimates $p_a(\mathbf{x}) = \frac{1}{\eta_a} p_i(\mathbf{x}) p_k(\mathbf{x})$ and $p_b(\mathbf{x}) = \frac{1}{\eta_b} p_j(\mathbf{x}) p_k(\mathbf{x})$. We fuse these dependent components according to the GMD.

$$\begin{aligned} p_c(\mathbf{x}) &= \frac{1}{\eta_c} p_a(\mathbf{x})^{\omega} p_b(\mathbf{x})^{1-\omega} \\ &= \frac{1}{\eta_c} p_i(\mathbf{x})^{\omega} p_j(\mathbf{x})^{1-\omega} p_k(\mathbf{x}). \end{aligned} \tag{26}$$

We suggest that if

$$H(p_r(\mathbf{x})) < \min\{H(p_a(\mathbf{x})), H(p_b(\mathbf{x}))\}, \tag{27}$$

then there exists a $\omega \in [0, 1]$ such that the minimum entropy GMD is

$$H(p_r(\mathbf{x})) \leq H(p_c(\mathbf{x})) \leq \min\{H(p_a(\mathbf{x})), H(p_b(\mathbf{x}))\}. \tag{28}$$

Conversely, if

$$H(p_r(\mathbf{x})) \geq \min\{H(p_a(\mathbf{x})), H(p_b(\mathbf{x}))\}, \tag{29}$$

then the minimum entropy GMD is

$$H(p_r(\mathbf{x})) \geq H(p_c(\mathbf{x})) \geq \min\{H(p_a(\mathbf{x})), H(p_b(\mathbf{x}))\}. \tag{30}$$

■

Thus, we expect that if (27) is true, then the entropy of the GMD is bounded below by $H(p_r(\mathbf{x}))$. Otherwise, it is bounded below by $\min\{H(p_a(\mathbf{x})), H(p_b(\mathbf{x}))\}$ and the minimum entropy occurs at $\omega = 0$ or $\omega = 1$.

It is easy to demonstrate empirically that (29) can be true, as shown in Section III, and this possibility raises an interesting conundrum for the GMD. If optimal fusion can increase entropy, by what criteria should one choose ω? It is no longer clear that minimising entropy makes sense. Arguably a more general measure is to minimise the divergence, such as the Kullbach-Leibler divergence, of the conservative approximation from the optimal pdf [7]. Properly justifying our optimisation criteria for ω remains as future work.

[7]We performed fusion experiments on randomly generated discrete probabilities and Gaussian mixture pdfs. We have as yet been unable to find a contradiction to Conjecture 1.

C. The GMD and the AMD

The GMD is not the only fusion rule to possess the "no double counting" property. An alternative is the *arithmetic mean density* (AMD), also referred to in the statistical literature as linear opinion pools [5],

$$p_m(\mathbf{x}) = \omega p_a(\mathbf{x}) + (1-\omega) p_b(\mathbf{x}). \tag{31}$$

As with the GMD, the AMD approximates the independent information and counts the common information only once.

Theorem 7: The AMD does not double count information.

$$\begin{aligned} p_m(\mathbf{x}) &= \omega p(\mathbf{x}|\mathbf{Z}_a) + (1-\omega) p(\mathbf{x}|\mathbf{Z}_b) \\ &\propto (\omega p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x}) + (1-\omega) p(\mathbf{Z}_b \backslash \mathbf{Z}_a|\mathbf{x})) \\ &\quad . p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}) \end{aligned} \tag{32}$$

Proof: The data is presumed to be conditionally independent, such that $p(\mathbf{Z}_a|\mathbf{x}) = p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x}) p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x})$, and so

$$\begin{aligned} &\omega p(\mathbf{x}|\mathbf{Z}_a) + (1-\omega) p(\mathbf{x}|\mathbf{Z}_b) \\ &\quad \propto \omega p(\mathbf{Z}_a|\mathbf{x}) p(\mathbf{x}) + (1-\omega) p(\mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}) \\ &\quad = \omega p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x}) p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}) \\ &\qquad\quad + (1-\omega) p(\mathbf{Z}_b \backslash \mathbf{Z}_a|\mathbf{x}) p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}) \\ &\quad = [\omega p(\mathbf{Z}_a \backslash \mathbf{Z}_b|\mathbf{x}) + (1-\omega) p(\mathbf{Z}_b \backslash \mathbf{Z}_a|\mathbf{x})] \\ &\qquad\quad \cdot p(\mathbf{Z}_a \cap \mathbf{Z}_b|\mathbf{x}) p(\mathbf{x}) \end{aligned}$$

■

However, the AMD and the GMD are qualitatively different rules for combining information. The GMD has a form similar to the *product rule* and Bayes' theorem, and so is properly seen as a fusion rule for dependent information. It addresses dependence but requires that each component pdf is individually consistent (e.g., the GMD will be inconsistent if $p_a(\mathbf{x}_t) = 0$). The AMD, on the other hand, performs amalgamation not fusion, and generates a mixture model from the component pdfs. Its form is related to the *sum rule*, and as such it *marginalises* over the *set of hypotheses* given by the component pdfs, and so can incorporate both consistent and inconsistent hypotheses.

The GMD is potentially inconsistent if a single component is inconsistent. The AMD is conservative if even a single component is consistent.

While the AMD behaves as a valid conservative fusion rule, it does not possess the potential to gain information, as described in Section IV-B. The "fused" solution always lies between the component pdfs

$$\min\{p_a(\mathbf{x}), p_b(\mathbf{x})\} \leq p_m(\mathbf{x}) \leq \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}, \ \forall \mathbf{x}. \tag{33}$$

By comparison, the GMD is not "bounded above" by $\max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}$.

Theorem 8: The GMD is not bounded above; it may exceed the maximum of its component functions [12]. There exists an $\mathbf{x}$ such that

$$p_c(\mathbf{x}) \geq \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}. \tag{34}$$

Proof: Suppose that (34) is false, so that

$$p_c(\mathbf{x}) < \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}, \quad \forall \mathbf{x}. \tag{35}$$

This implies

$$\begin{aligned} f(\mathbf{x}, \omega) &= \eta_c p_c(\mathbf{x}) \\ &< \eta_c \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\} \\ &<= \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}, \quad \forall \mathbf{x}, \omega, \end{aligned} \tag{36}$$

In particular, for $\omega = 0$, we get $f(\mathbf{x}, \omega) = p_b(\mathbf{x}) < p_b(\mathbf{x})$, which is clearly false. By contradiction, there must exist at least one x such that $p_c(\mathbf{x}) \geq \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}$. ■

This "not bounded above" property is also supported by (21), since the value of $p_c(\mathbf{x})$ will equal or exceed $p_a(\mathbf{x})$ and $p_b(\mathbf{x})$ at every point where they intersect, whereas the AMD will merely give equality. We can further strengthen this property to show that whenever $\eta_c < 1$ the GMD result will strictly exceed both component pdfs at some point.

Corollary 1: The GMD will strictly exceed the component functions for some $\mathbf{x}$ if and only if $\eta_c < 1$.

$$\exists \mathbf{x} : p_c(\mathbf{x}) > \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\} \iff \eta_c < 1 \tag{37}$$

Proof: Similar to Theorem 8, we obtain a contradiction to the supposition that $p_c(\mathbf{x}) \leq \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}$ for all $\mathbf{x}$ when $\eta_c < 1$, since this implies $f(\mathbf{x}, \omega) \leq \eta_c \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\} < \max\{p_a(\mathbf{x}), p_b(\mathbf{x})\}$. ■

V. Conclusion

This paper investigates the notions of consistency and conservativeness in the context of Bayesian data fusion. We consider existing justifications of Bayes' theorem; it has "structural" consistency, generates unique solutions, preserves information, and does not double count common information. We propose that a conservative estimate is one that is consistent when dependencies between component pdfs are neglected or unknown.

The GMD is justified as a conservative fusion rule. It has the same essential "product form" as Bayes' theorem, it does not double count common information, and it replaces the independent information with a conservative approximation. A consequence of this form is that GMD estimates are "bounded below" point-wise by the minimum of the component pdfs. This prevents inconsistent behaviour when applied to *recursive* fusion with dependent information. The GMD also has the potential to reduce entropy, indicating its ability to effectively combine available information.

Further to providing justification of the GMD as a conservative fusion rule (and the CI algorithm as a special case), this paper provides general insights into the nature of conservativeness. We see not double counting common information as a necessary condition for consistency. We demonstrate that entropy is not a sufficient condition for conservativeness, and suggest that conservative approximations involve both entropy increase and order preservation. A surprising property of non-Gaussian data fusion is that entropy may increase, which shows that fundamentally Bayesian data fusion concerns combining information, not reducing uncertainty. The possibility of entropy-increasing fusion raises a question for the GMD as to the optimisation of ω. We shall investigate the notion of optimality for conservative estimation in future work.

Acknowledgments

The authors would like to thank Sisir Karumanchi, Ian Mahon, Paul Thompson and Hugh Durrant-Whyte for valuable discussions.

Appendix

The following proof of Lemma 1 shows that the derivative of the differential entropy with respect to ω is non-positive. This, in turn, implies that the entropy is non-increasing for $\omega > 0$ and, in particular, that $H(p(\mathbf{x})) \leq H(p_\omega(\mathbf{x}))$ for $0 < \omega \leq 1$, where $p_\omega(\mathbf{x}) = \frac{1}{\eta_\omega} p(\mathbf{x})^\omega$ and $\eta_\omega = \int p(\mathbf{x})^\omega \, d\mathbf{x}$.

We first note that

$$\ln p_\omega(\mathbf{x}) = \omega \ln p(\mathbf{x}) - \ln \eta_\omega, \tag{38}$$

and, given the identity $\frac{du^x}{dx} = u^x \ln u$, that

$$\frac{d\eta_\omega}{d\omega} = \int \frac{d}{d\omega} p(\mathbf{x})^\omega \, d\mathbf{x} = \int p(\mathbf{x})^\omega \ln p(\mathbf{x}) \, d\mathbf{x}. \tag{39}$$

Therefore,

$$\frac{d \ln p_\omega(\mathbf{x})}{d\omega} = \ln p(\mathbf{x}) - \frac{1}{\eta_\omega} \int p(\mathbf{x})^\omega \ln p(\mathbf{x}) \, d\mathbf{x}. \tag{40}$$

Given the identity $\frac{du}{dx} = u \frac{d \ln u}{dx}$, we get

$$\begin{aligned} \frac{dp_\omega(\mathbf{x})}{d\omega} &= \frac{p(\mathbf{x})^\omega \ln p(\mathbf{x})}{\eta_\omega} - \frac{p(\mathbf{x})^\omega}{\eta_\omega^2} \int p(\mathbf{x})^\omega \ln p(\mathbf{x}) \, d\mathbf{x} \\ &= p_\omega(\mathbf{x}) \ln p(\mathbf{x}) - p_\omega(\mathbf{x}) \int p_\omega(\mathbf{x}) \ln p(\mathbf{x}) \, d\mathbf{x}. \end{aligned} \tag{41}$$

Furthermore,

$$\begin{aligned} \int \frac{dp_\omega(\mathbf{x})}{d\omega} d\mathbf{x} &= \int p_\omega(\mathbf{x}) \ln p(\mathbf{x}) \, d\mathbf{x} \\ &\quad - \underbrace{\int p_\omega(\mathbf{x}) \, d\mathbf{x}}_{=1} \int p_\omega(\mathbf{x}) \ln p(\mathbf{x}) \, d\mathbf{x} \\ &= 0. \end{aligned} \tag{42}$$

Therefore, the derivative of the differential entropy with respect to ω is as follows

$$\begin{aligned}\frac{dH\left(p_{\omega}\left(\mathbf{x}\right)\right)}{d\omega} &= -\int \frac{d}{d\omega}(p_{\omega}\left(\mathbf{x}\right)\ln p_{\omega}\left(\mathbf{x}\right)d\mathbf{x}) \\ &= -\int \frac{dp_{\omega}\left(\mathbf{x}\right)}{d\omega}\ln p_{\omega}\left(\mathbf{x}\right)d\mathbf{x} - \underbrace{\int \frac{dp_{\omega}\left(\mathbf{x}\right)}{d\omega}d\mathbf{x}}_{=0}.\end{aligned} \tag{43}$$

Expanding (43) in terms of to (41), we get

$$\begin{aligned}\frac{dH\left(p_{\omega}\left(\mathbf{x}\right)\right)}{d\omega} &= -\underbrace{\int p_{\omega}\left(\mathbf{x}\right)\ln p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}}_{(a)} \\ &+ \underbrace{\int p_{\omega}\left(\mathbf{x}\right)\ln p_{\omega}\left(\mathbf{x}\right)d\mathbf{x}\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}}_{(b)}.\end{aligned} \tag{44}$$

The component terms may be further expanded as

$$\begin{aligned}(a) &= \int p_{\omega}\left(\mathbf{x}\right)\left(\omega\ln p\left(\mathbf{x}\right) - \ln\eta_{\omega}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x} \\ &= \omega\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)^{2}d\mathbf{x} - \ln\eta_{\omega}\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x},\end{aligned} \tag{45}$$

$$\begin{aligned}(b) &= \int p_{\omega}\left(\mathbf{x}\right)\left(\omega\ln p\left(\mathbf{x}\right) - \ln\eta_{\omega}\right)d\mathbf{x}\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x} \\ &= \left(\omega\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x} - \ln\eta_{\omega}\right)\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}.\end{aligned} \tag{46}$$

Substituting these back into (44) we get

$$\begin{aligned}&\frac{dH\left(p_{\omega}\left(\mathbf{x}\right)\right)}{d\omega} \\ &= -\omega\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)^{2}d\mathbf{x} + \omega\left(\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}\right)^{2} \\ &= -\omega\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)^{2}d\mathbf{x} \\ &\qquad + 2\omega\left(\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}\right)^{2} \\ &\qquad - \omega\left(\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}\right)^{2}\underbrace{\int p_{\omega}\left(\mathbf{x}\right)d\mathbf{x}}_{=1} \\ &= -\omega\int p_{\omega}\left(\mathbf{x}\right)\left[\ln p\left(\mathbf{x}\right)^{2} - 2\ln p\left(\mathbf{x}\right)\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}\right. \\ &\qquad\left. + \left(\int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}\right)^{2}\right]d\mathbf{x} \\ &= -\omega\int p_{\omega}\left(\mathbf{x}\right)\left(\ln p\left(\mathbf{x}\right) - \int p_{\omega}\left(\mathbf{x}\right)\ln p\left(\mathbf{x}\right)d\mathbf{x}\right)^{2}d\mathbf{x}.\end{aligned} \tag{47}$$

This is clearly non-positive for $\omega > 0$.

REFERENCES

[1] T. Bailey, M. Bryson, H. Mu, J. Vial, L. McCalman, and H. Durrant-Whyte. Decentralised cooperative localisation for heterogeneous teams of mobile robots. In *IEEE International Conference on Robotics and Automation*, 2011.

[2] C.Y. Chong, S. Mori, and K.C. Chang. Distributed multitarget multisensor tracking. In Y. Bar-Shalom, editor, *Multitarget-Multisensor Tracking: Advanced Applications*, pages 247–295. Artech House, 1990.

[3] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[4] A.G. Dabak. *A geometry for detection theory*. PhD thesis, Rice University, Department of Electrical and Computer Engineering, 1992.

[5] C. Genest and J.V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–148, 1986.

[6] T. Heskes. Selecting weighting factors in logarithmic opinion pools. In *Advances in Neural Information Processing Systems*, pages 266–272, 1998.

[7] M. Hurley. An information-theoretic justification for covariance intersection and its generalization. In *IEEE International Conference on Information Fusion*, 2002.

[8] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[9] S. Julier. Fusion without independence. In *IET Seminar on Tracking and Data Fusion: Algorithms and Applications*, pages 1–5, 2008.

[10] S.J. Julier. An empirical study into the use of Chernoff information for robust, distributed fusion of Gaussian mixture models. In *IEEE International Conference on Information Fusion*, 2006.

[11] S.J. Julier. Estimating and exploiting the degree of independent information in distributed data fusion. In *IEEE International Conference on Information Fusion*, 2009.

[12] S.J. Julier, T. Bailey, and J.K. Uhlmann. Using exponential mixture densities for suboptimal distributed data fusion. In *IEEE Nonlinear Statistical Signal Processing Workshop*, 2006.

[13] S.J. Julier and J.K. Uhlmann. A non-divergent estimation algorithm in the presence of unknown correlations. In *American Control Conference*, volume 4, pages 2369–2373, 1997.

[14] D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[15] R. Mahler. Optimal/robust distributed data fusion: A unified approach. In *SPIE International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, volume 4052, pages 128–138, 2000.

[16] J. O'Brien. An algorithm for the fusion of correlated probabilities. In *IEEE International Conference on Information Fusion*, pages 565–571, 1998.

[17] P. Thompson. General conservative fusion is conservative. Technical report, University of Sydney, Australian Centre for Field Robotics, 2006.

[18] S.W. Utete. *Network management in decentralised sensing systems*. PhD thesis, University of Oxford, Department of Engineering Science, 1995.

[19] A. Zellner. Optimal information processing and Bayes theorem. *American Statistician*, 42(4):278–280, 1988.

[20] A. Zellner. Information processing and Bayesian analysis. *Journal of Econometrics*, 107:41–50, 2002.