Abstract—The paper deals with the merging of hypotheses that are not provided with weights and are represented by probability densities. A recently proposed definition of a conservative probability density is exploited to evolve the ideas of the covariance union approach. It is derived that the solution with the lowest entropy is given by the mixture density with the maximum entropy and a closed form solution for disjoint supports is presented. The proposed approach is also applicable to discrete random variables. The paper is concluded by illustrative examples.

Index Terms—conservative density, generalised covariance union, mixture density, data fusion, nonlinear estimation

I. INTRODUCTION

There are several features that an estimator should not have. An example is overconfidence, because underestimating the uncertainty may be hazardous. On the other hand, being cautious, i.e. conservative, is costly. Nevertheless, conservativeness is usually considered to be the lesser evil.

In target tracking [1], multiple-hypothesis tracking arises from many problem formulations [2]–[6], for example from tracking in a clutter, multitarget tracking, multisensor tracking, multimodel tracking, ground tracking within a road network or from simultaneous localisation and map building. The common issue is a perpetual growth of the number of hypotheses. Since computation resources are always limited, a pruning or merging of the hypotheses has to be applied.

A distinctive feature of the approaches to the hypothesis-number reduction is whether the hypotheses are provided with weights or all the hypotheses are stated possible with no further preference. It is useful to remark that the unweighted case is not identical to the weighted case with all weights being equal. Whereas the weighted hypotheses can be considered to be the components of a mixture density, the unweighted hypotheses do not permit such interpretation. There are many approaches to the mixture reduction [7], for example the pruning of the components with negligible weights, successive merging of the nearest components or minimisation of a global distance criterion. However, there are few approaches regarding the unweighted hypotheses which will be elaborated in this paper.

The main obstacle to the development of approaches which consider the unweighted hypotheses is the lack of a definition of a specific partial order over the hypotheses or at least a definition of a specific reflexive antisymmetric binary relation. In other words, it is not known how to appropriately compare two hypotheses, ideally in the transitive way. If all the hypotheses were in the relation with a unique density, that may be given by a mixture density, it would be possible to replace the set of hypotheses with the density, and therefore a reduction of the number of hypotheses would be enabled.

So far, the relation has been defined for the hypotheses given by the estimates in the form of the mean value and an associated covariance matrix of the estimate error. If the estimated covariance matrix is not lower than the mean square error matrix, i.e. if the estimated covariance matrix is not lower than the true covariance matrix increased by a square of the bias of the mean value, then the estimate is conservative.

This definition of a conservative estimate induced the covariance union approach [8]–[10]. A set of estimates with no associated weights can be replaced by an estimate that is conservative with respect to each estimate in the set. The inherent flaw, inherited from the definition, is the lack of the transitivity. Thus an estimate which is conservative with respect to a conservative estimate need not be conservative. It is possible to ensure conservativeness, but the price is high and seldom pays off.

The covariance union approach has been generalised in [11]. Confining the possible values of the weights to intervals connects the cases of unweighted and weighted estimates. The unit interval represents the case of no weights, the degenerate intervals corresponds to the case of given weights. However, the generalisation does not concern the estimates in the form of the probability densities, it does not concern dependent hypotheses. If several hypotheses form a mixture density, the generalised approach does not treat them en bloc.

Recently, the authors have defined the conservativeness of estimates in the form of probability densities [12]. The definition has been derived from the definition for the mean value and covariance matrix and both definitions are equivalent for one dimensional Gaussian densities. Of course, this intentional equivalence leads to inheriting the transitivity issue, a density which is conservative with respect to a conservative density need not be conservative.

Provided with the definition of a conservative density, the aim of the paper is to design a density that is conservative with respect to all merged densities but that is not conservative unnecessarily. Such an approach is analogous to the covariance union approach, however, it is considered for probability
densities. Beside the merging of composite hypotheses in the above mentioned tracking problems, the approach can be also used in fusion problems [13], [14].

The paper is organised as follows. The problem is stated in Section II. The derivation of the conservative density is given in Section III, an explicit solution for a special case is provided in Section IV. Section V offers a further discussion. Illustrative examples are given in Section VI. The main results are summarised in Section VII.

II. PROBLEM STATEMENT

Prior stating the problem, the definition of a conservative estimate and the covariance union approach are reminded.

Let the system state $x$ be given by a vector random variable. A state estimate which is given by $(\mu_x, P_\pi)$, where $\mu_x$ is the estimated mean and $P_\pi$ is the estimated covariance matrix, is conservative if the following condition holds,

$$P_\pi - P_1 - (\mu_1 - \mu_x)(\mu_1 - \mu_x)^T \succeq 0. \quad (1)$$

The inequality $\succeq 0$ means that the left hand side term is positive semidefinite.

The covariance union approach assumes that there are $N$ estimates $(\mu_x, P_i), i = 1, \ldots, N,$ that may be conditioned by different data or be spurious due to an ambiguous association for example. The goal is to find and estimate $(\mu_\pi, P_\pi)$ that is conservative with respect to each merged estimate, i.e. it has to obey $N$ matrix inequalities

$$P_\pi - P_i - (\mu_i - \mu_x)(\mu_i - \mu_x)^T \succeq 0. \quad (2)$$

Since there are many such estimates, the determinant of the estimated covariance matrix $P_\pi$ is chosen as the criterion to be minimised.

Let the probability density of the state be denoted $p$. According to [12], the density $\pi$ is conservative, if the following condition holds,

$$\mathcal{H}(\pi) - \mathcal{H}(p) - \mathcal{D}(p\|\pi) \geq 0, \quad (3)$$

where $\mathcal{H}(\pi)$ refers to the (differential) Shannon entropy,

$$\mathcal{H}(\pi) = -\int_{\Omega_\pi} \pi(x) \ln \pi(x) \, dx, \quad (4)$$

$\Omega_\pi = \{x : \pi(x) > 0\}$ is the support of $\pi$, and $\mathcal{D}(p\|\pi)$ denotes the Kullback–Leibler divergence,

$$\mathcal{D}(p\|\pi) = \int_{\Omega_p} p(x) \ln \frac{p(x)}{\pi(x)} \, dx \quad (5)$$

for $p$ absolutely continuous with respect to $\pi$, $p \ll \pi$, which means that $p(x) > 0 \Rightarrow \pi(x) > 0$. For $p$ not absolutely continuous with respect to $\pi$, the divergence $\mathcal{D}(p\|\pi)$ is infinite. The left hand side term of (3) will be denoted by $\mathcal{C}(p\|\pi)$. Note that comparing to the condition of conservative estimate (1), the condition of conservative density (3) is scalar.

Now, suppose that densities $p_i, i = 1, \ldots, N,$ are given. The goal is to find a density $\pi$ that is conservative with respect to the densities $p_i$, i.e. $\mathcal{C}(p_i\|\pi) \geq 0$. In order not to introduce a redundant uncertainty, the entropy $\mathcal{H}(\pi)$ should be minimised.

III. THE MAXIMUM ENTROPY MIXTURE IS CONSERVATIVE

According to the definitions of the conservativeness (3), entropy (4) and divergence (5), the conditions $\mathcal{C}(p_i\|\pi) \geq 0$ can be rewritten as

$$\int_{\Omega_\pi} (p_i(x) - \pi(x)) \ln \pi(x) \, dx \geq 0, \quad (6)$$

where it must hold $p_i \ll \pi$ and therefore $\Omega_{p_i} \subseteq \Omega_\pi$.

Suppose that the conservative density $\pi$ is known. Now, a necessary condition will be derived. For an arbitrary choice of weights $\omega_i, \omega_i \geq 0, \sum_{i=1}^N \omega_i = 1,$ a density $\pi$ fulfilling the conditions (6) necessarily fulfils the condition

$$\int_{\Omega_\pi} \left( \sum_{i=1}^N \omega_i p_i(x) - \pi(x) \right) \ln \pi(x) \, dx \geq 0, \quad (7)$$

since the sum of nonnegative numbers $\omega_i \mathcal{C}(p_i\|\pi)$ is nonnegative – multiply (6) by $\omega_i$ for each $i$, sum the inequalities, switch the summation and integration, factorise the logarithm and consequently the density $\pi$ and substitute the sum of weights by one. Rewriting the condition (7) to the entropy and divergence terms gives the condition

$$\mathcal{H}(\pi) - \mathcal{H}(\sum_{i=1}^N \omega_i p_i) - \mathcal{D}(\sum_{i=1}^N \omega_i p_i\|\pi) \geq 0. \quad (8)$$

In other words, a density $\pi$ conservative with respect to the densities $p_i$ is necessarily conservative with respect to all mixtures of these densities. Especially, the density $\pi$ is conservative with respect to the mixture density with the highest entropy. This mixture will be denoted by $\pi^*$. Now, the derived necessary condition (7) can be exploited. The Kullback–Leibler divergence (5) is positive for $\pi \neq p$ and zero for $\pi = p$, where the equality holds up to a zero measure set. Therefore, the entropy $\mathcal{H}(\pi)$ of a conservative density $\pi$ cannot be lower than the maximum mixture entropy $\mathcal{H}(\pi^*)$, see (3). A conservative density $\pi$ with the same entropy, $\mathcal{H}(\pi) = \mathcal{H}(\pi^*)$, cannot differ from the density $\pi^*$ over a set with nonzero measure, because the density $\pi$ has to be conservative with respect to the density $\pi^*$, i.e. to the mixture with the maximum entropy.

Recall that in order not to introduce redundant uncertainty, the density with the minimum entropy is searched. So if the maximum entropy mixture $\pi^*$ is conservative with respect to all components, the solution has just been found. Denote the set of all mixtures of the densities $p_i$ by $\mathcal{M}(p_i)$. This set is convex, $p_A \in \mathcal{M}(p_1) \land p_B \in \mathcal{M}(p_1) \Rightarrow \omega p_A + (1 - \omega) p_B \in \mathcal{M}(p_1)$ for $0 \leq \omega \leq 1$.

Further, the uniform density $p_U$ on the support $\Omega_\pi = \bigcup_{i=1}^N \Omega_{p_i}$ is given by $p_U(x) = |\Omega_\pi|^{-1}, x \in \Omega_\pi$, where $|\Omega_\pi|$ means the measure of the support $\Omega_\pi$. The uniform density $p_U$ is universally conservative. It is conservative with respect to all densities $p$ with the same or smaller support, $P_p \subseteq \Omega_{p_u}$, that is straightforward from (6). The logarithm of a constant can be factorised out and the integral over a density is 1 for both densities. However, the conservativeness $\mathcal{C}(p\|p_u)$ is always zero.
If the universally conservative density \( p_u \) can be expressed as the mixture of the merged densities \( p_i \), i.e. \( p_u \in \mathcal{M}(p_i) \), it is the solution itself, \( \pi^* = p_u \).

In the opposite case, the “Pythagorean” theorem 11.6.1 in [15] can be exploited with the substitution \( Q \rightarrow p_u, P \rightarrow p, p \in \mathcal{M}(p_i) \), and \( P^* \rightarrow \pi^* \). It it also needed to recall that \( \mathcal{D}(p\|p_u) = \ln \Omega_{p_u} - \mathcal{H}(p) \). According to the theorem, the maximum entropy mixture \( \pi^* \) is conservative with respect to each mixture \( p \in \mathcal{M}(p_i) \).

The proof is the following. Consider a mixture \( \omega p + (1-\omega)\pi^* \) and compute the derivative of its entropy with respect to \( \omega \) at \( \omega = 0 \). As the entropy of \( \pi^* \) reaches its maximum on a convex set \( \mathcal{M}(p_i) \), the derivative is non-positive. Thus, the conservativeness condition has just been obtained,

\[
0 \geq \frac{d}{d\omega} \mathcal{H}(\omega p + (1-\omega)\pi^*) \bigg|_{\omega=0} = -\int_{\Omega_{\pi^*}} (p - \pi^*) + (p - \pi^*) \ln(\omega p + (1-\omega)\pi^*) \omega=0 \, d\mathbf{x} = -\mathcal{C}(p\|\pi^*). \tag{9}
\]

This section is summarised by its title, the maximum entropy mixture \( \pi^* \) is the solution to the given problem. It is conservative with respect to all the components \( p_i \). Moreover, it is conservative with respect to all possible mixtures \( p, p \in \mathcal{M}(p_i) \).

The problem of searching the function that fulfils the requirements has been transformed into the problem of searching the optimal parameters. Various optimisation algorithms can be used [16], but they will not be treated in this paper. Note also that the entropy is a concave function [15].

IV. EXPICIT SOLUTION FOR DISJOINT SUPPORTS

In the special case that the supports of the densities \( p_i \) are disjoint, \( \Omega_{p_i} \cap \Omega_{p_j} = \emptyset, i \neq j \), the component weights \( \omega^*_i \) of the maximum entropy mixture \( \pi^* \) can be obtained in a closed form.

Since the condition \( p_i \ll \pi^* \) must hold in order to enable (6), the weights \( \omega^*_i \) cannot be zero. Therefore, the support of the maximum entropy mixture \( \pi^* \) is given by the union of the supports, \( \Omega_{\pi^*} = \bigcup_{i=1}^{N} \Omega_{p_i} \).

Consider the necessary condition (7), where the maximum entropy mixture and the corresponding weights are used, i.e. the substitution \( \pi \rightarrow \pi^*, \omega_i \rightarrow \omega^*_i \) is made. Recall that (7) has been obtained as a weighted sum of (6). Further, the left hand side of (7) is now equal to \( \mathcal{C}(\pi^*\|\pi^*) \) which is zero according to the definition (3). In general, if the sum of nonnegative numbers is zero, then all the summands have to be zero. Since the weights \( \omega^*_i \) are positive, the equality must hold in all conditions (6), i.e. \( \mathcal{C}(p_i\|\pi^*) = 0 \).

According to the disjointness of the supports, it holds \( \pi^*(\mathbf{x})|_{\mathbf{x} \in \Omega_{p_i}} = \omega^*_i p_i(\mathbf{x}) \). The derived conditions \( \mathcal{C}(p_i\|\pi^*) = 0 \) can be rewritten as

\[
\int_{\Omega_{p_i}} (1 - \omega^*_i)p_i(\mathbf{x}) \ln(\omega^*_i p_i(\mathbf{x})) \, d\mathbf{x} = \sum_{j=1, \ldots, N, j \neq i} \int_{\Omega_{p_j}} (1 + \omega^*_j p_j(\mathbf{x})) \ln(\omega^*_j p_j(\mathbf{x})) \, d\mathbf{x} = 0. \tag{10}
\]

Since \( 1 - \omega^*_i = \sum_{j=1, \ldots, N, j \neq i} \omega^*_j \) and since the logarithm of a product is equal to the sum of the logarithms of the factors and the densities integrate to one, these equations can be rewritten further as

\[
\sum_{j=1, \ldots, N, j \neq i} \omega^*_j [\mathcal{H}(p_i) + \ln \omega^*_i + \mathcal{H}(p_j) - \ln \omega^*_j] = 0. \tag{11}
\]

According to the previous discussion, the weights \( \omega^*_i \) have to be positive. The obtained set of nonlinear equations (11) can be solved by setting the brackets to zero. This gives the following equations,

\[
\frac{\omega^*_i}{\omega^*_j} = \frac{e^{\mathcal{H}(p_i)}}{e^{\mathcal{H}(p_j)}}. \tag{12}
\]

The weights \( \omega^*_i \) of the maximum entropy mixture have been expressed in a closed form. The remaining inconvenience is that the entropy (5) has an analytic solution only in special cases and has to be computed numerically or estimated [17] in a general case.

V. FURTHER INSIGHTS

This section provides a further insight into the geometry of the conservativeness. Standard optimisation techniques will be used, see for example [16], [18].

Consider mixture densities \( \pi = \sum_{i=1}^{N} \omega_i p_i \), i.e. \( \pi \in \mathcal{M}(p_i) \).

The minimisation of the criterion given by the negative entropy,

\[
-\mathcal{H}(\pi) = \int_{\Omega_{\pi}} \left( \sum_{i=1}^{N} \omega_i p_i(\mathbf{x}) \ln(\sum_{i=1}^{N} \omega_i p_i(\mathbf{x})) \right) d\mathbf{x}, \tag{14}
\]

subject to the equality and inequality constraints, \( \sum_{i=1}^{N} \omega_i = 1, -\omega_i \leq 0 \), will be treated.

The optimal weights \( \omega^*_i \) have to obey the Karush–Kuhn–Tucker conditions. For each \( i = 1, \ldots, N \), it has to hold

\[
0 = \left\{ \frac{d}{d\omega_i} [-\mathcal{H}(\pi) + \lambda(\sum_{j=1}^{N} \omega_j - 1) + \sum_{j=1}^{N} \mu_j (-\omega_j)] \bigg|_{\omega^*_i} \right\} = \int_{\Omega_{\pi}} (p_i \ln(\sum_{j=1}^{N} \omega^*_j p_j) + p_i) \, d\mathbf{x} + \lambda - \mu_i, \tag{15}
\]

\[
0 = \mu_i (-\omega^*_i), \quad 0 \leq \mu_i, \tag{16}
\]

where \( \mu_i \) are the multipliers corresponding to the inequality constraints on the weights \( \omega_i \), the multiplier \( \lambda \) corresponds to the equality constraint, and \( \omega^*_i = \omega^*_j \) for \( j = 1, \ldots, N \).
Multiplying the equations (15) by $\omega_i^*$ and summing those corresponding to the inactive inequality constraints, i.e. $i \in \mathcal{M}$, $\mathcal{M} = \{ j : \mu_j = 0 \}$, the following equation is obtained,

$$\int_{\Omega_\pi} \left( \sum_{i \in \mathcal{M}} \omega_i^* p_i \right) \ln \left( \sum_{j=1}^{N} \omega_j^* p_j \right) dx + \sum_{i \in \mathcal{M}} \omega_i^* (1 + \lambda) = 0, \quad (18)$$

For the active inequality constraints, $k \in \{1, \ldots, N\} \setminus \mathcal{M}$, the multipliers $\mu_k$ are nonzero, and thus the corresponding weights $\omega_k^*$ have to be zero, see (16). This leads to $\sum_{i \in \mathcal{M}} \omega_i^* = 1$ and $\sum_{i \in \mathcal{M}} \omega_i^* p_i = \sum_{j=1}^{N} \omega_j^* p_j = \pi^*$, i.e. the maximum entropy mixture $\pi^*$ need not be given by all components $p_j$, $j = 1, \ldots, N$, but only by several $p_i$, $i \in \mathcal{M}$.

By subtracting the equation (18) from the equations (15), the multiplier $\lambda$ is cancelled out and the following conditions are obtained for $i = 1, \ldots, N$,

$$\int_{\Omega_\pi} (p_i - \pi^*) \ln \pi^* dx - \mu_i = 0. \quad (19)$$

Thus, the conservativeness of the maximum entropy mixture $\pi^*$ with respect to its components $p_i$, $i \in \{ j : \omega_j^* > 0 \}$, is zero, since the corresponding multiplier $\mu_i$ has to be zero, see (16). Further, the conservativeness $C(p_k || \pi^*)$ is nonnegative for the other densities $p_k$, $k \in \{ j : \omega_j^* = 0 \}$, since the corresponding multiplier $\mu_k$ has to be nonnegative, see (17). Moreover, due to the linearity in the first argument, the conservativeness $C(p_k || \pi^*)$ is positive if and only if the component $p_k$ is linearly independent of the components $p_i$, $i \in \{ j : \omega_j^* > 0 \}$.

The Karush–Kuhn–Tucker conditions have offered another proof that the maximum entropy mixture $\pi^*$ is conservative with respect to the individual hypotheses that are given by the densities $p_i$, $i = 1, \ldots, N$. In addition, the conditions have also shown in which cases the conservativeness of $\pi^*$ with respect to $p_i$ is positive.

Another remark on the conservativeness is that it is possible to define a universally conservative density $p_c$ other than the uniform density $p_u$ over the united support $\Omega_\pi$. This may be useful if a regular nonlinear transformation of coordinates is considered. The uniform density in the Cartesian coordinates does not transform to the uniform density in the polar coordinates. In polar coordinates, it can be reasonable to choose the transformed uniform density to be the universally conservative density.

The modification of the definition (3) is the following condition,

$$D(p || p_c) - D(\pi || p_c) - D(p || \pi) \geq 0, \quad (20)$$

and the counterpart to the condition (6) is the condition

$$\int_{\Omega_\pi} (p_i(x) - \pi(x)) \ln \frac{\pi(x)}{p_c(x)} dx \geq 0. \quad (21)$$

Straightforward modifications of the equations derived up to now have to be done. For example, the optimal weights (13) are newly given by $\omega_i^* \propto e^{-D(p_i || p_c)}$, where $\propto$ means proportional to, and $D(\pi || p_c)$ is minimised instead of the negative entropy in (14). The nearest mixture is searched instead of the maximum entropy mixture and so on.

### VI. Illustrative Examples

This section gives illustrative examples. The explicit solution for special densities is treated first. A balancing property of the conservative merging is pointed out. Next, the applicability of the proposed approach to discrete random variables is dealt with. The random variables with three possible states offer nice geometrical insights. The merging of Gaussian hypotheses is inspected last.

#### A. Disjoint affine densities

Let the densities $p_i$, $i = 1, \ldots, N$ be uniform densities with disjoint supports, $\Omega_{p_i} \cap \Omega_{p_j} = \emptyset$, $i \neq j$, $p_i(x) = |\Omega_{p_i}|^{-1} \alpha_i 1$. The corresponding entropies are $H(p_i) = \ln |\Omega_{p_i}|$. Hence, due to (13), the weights $\omega_i^*$ of the maximum entropy mixture $\pi^*$ are proportional to $|\Omega_{p_i}|$. Thus, the maximum entropy mixture is given by

$$\pi^*(x) = \left\{ \sum_{i=1}^{N} |\Omega_{p_i}|^{-1} \alpha_i 1, \quad x \in \Omega_\pi = \bigcup_{i=1}^{N} \Omega_{p_i}, \right\} \quad (22)$$

i.e. by the uniform density over the united support.

This result has been expected. The uniform density is universally conservative and can be obtained as a mixture of the uniform densities over the partial supports.

Let the densities $p_i$ be Gaussian with the mean vectors $\mu_i$ and covariance matrices $P_i$, $p_i = \mathcal{N}(\mu_i, P_i)$. Suppose the densities are far away from each other in the sense that for $i \neq j$ it holds $(\mu_i - \mu_j)^T (P_i + P_j)^{-1} (\mu_i - \mu_j) \gg 1$. By neglecting the tails, the supports can be considered disjoint and an approximation $\pi^*$ of the maximum entropy mixture $\pi^*$ can be obtained.

The entropy of the Gaussian densities is given by $H(p_i) = \frac{1}{2} \ln \left\{ (2\pi e)^{n_x} |P_i| \right\}$, where $n_x$ is the dimension of $x$ and $|P|$ denotes the determinant of $P$. The weights $\omega_i^*$ are obtained by (13) as

$$\omega_i^* \propto \sqrt{\det P} \sqrt{(2\pi e)^{n_x} |P_i|}.$$  

The second square root term is the normalising constant of the corresponding Gaussian density. Therefore, all exponential terms of the resulting Gaussian mixture will have a common normalising constant.

The approximation of the maximum entropy mixture is given by

$$\pi^*(x) = \left\{ \sum_{i=1}^{N} |P_i|^{\frac{1}{2}} p_i(x) \right\} \left\{ \sum_{j=1}^{N} |P_j|^{\frac{1}{2}} \right\}^{-1}. \quad (23)$$

The density at all the means $\mu_i$ has approximately the same value, $\pi^*(\mu_i) \approx (2\pi e)^{-\frac{n_x}{2}} (\sum_{j=1}^{N} |P_j|^{\frac{1}{2}})^{-1}$, so a kind of uniformity can be observed.

Fig. 1 depicts the following example. The densities $p_i$, $i = 1, 2, 3$, for a scalar state $x$, are given by

$$p_1 = \mathcal{N}(-17, 4), \quad p_2 = \mathcal{N}(-2, 1), \quad p_3 = \mathcal{N}(15, 9), \quad (24)$$

and the maximum entropy mixture can be approximated by the following mixture,

$$\pi^*(x) = \frac{1}{3} p_1(x) + \frac{1}{6} p_2(x) + \frac{1}{2} p_3(x). \quad (25)$$
It can be observed that the conservative merging harmonises the resolution in the sense that the probability per volume is being equalised across the components. The exponential function of the entropy gives the effective volume of the support [15]. Thus, the components with large effective support volumes are weighted by large weights. If one wants to quantise a continuous random variable, more discrete points are needed for representing larger volumes to achieve a given quantisation error. Therefore, it is desirable that the merging allocates more points to the components with large effective volumes of the support. As the proportion of the allocated points is related to the component weight, the conservative merging of disjoint densities balances the errors across the components.

The balancing property can be generalised. Consider general densities \( p_i \) that are the same up to a different position and scaling parameters \( b_i, A_i \). If there exist a density \( p \) such that \( p_i(x) = p(A_i^{-1}(x - b_i))(\text{abs } |A_i|^{-1}) \), where \( \text{abs} \) denotes the absolute value, the entropies are related by \( H(p_i) = H(p) + \ln \text{abs } |A_i| \). Further, if the supports are disjoint, the maximum entropy mixture is given by

\[
\pi^*(x) = \{ \sum_{i=1}^{N} p(A_i^{-1}(x - b_i)) \} \{ \sum_{j=1}^{N} \text{abs } |A_j| \}^{-1}. \tag{26}
\]

Again, the supports of the densities are scaled individually, but the normalisation is applied at once.

**B. Piecewise constant densities**

It is easy to show that the conservativeness can also be considered in the discrete case.

Let us have a piecewise constant density \( p(x) \) such that

\[
p(x) = \sum_{k=1}^{M} P_k 1_{\Omega_k}(x), \tag{27}
\]

where \( 1_{\Omega_k} \) is the indicator function of the support of the cell \( k \), the cells have the same volume, \( |\Omega_k| \propto \frac{1}{2^M} \), and are disjoint, \( \Omega_k \not= \Omega_l, k \not= l \). Without a loss of generality, the unit volume of the cells can be considered, \( |\Omega_k| = 1 \), since a regular linear transformation preserves the conservativeness [12]. By solving the integrals, the differential entropy (4) and the Kullback–Leibler divergence (5) convert to their discrete counterparts.

The proposed approach will be demonstrated on discrete random variables with three states, \( M = 3 \). The probability functions \( p \) will be given by vectors of the probabilities, \( p = [P_1, P_2, P_3]^T \). The components \( p_i \) are given by \( p_1 = [1, 0, 0]^T \), \( p_2 = [0.5, 0.5, 0]^T \), \( p_3 = [0.3, 0.6, 0.1]^T \), \( p_4 = [0.6, 0.1, 0.3]^T \) and \( p_5 = [0.2, 0.2, 0.6]^T \).

Fig. 2 and 3 illustrate special constellations of the components. The zero conservativeness contours partition the probability simplex \( \mathcal{M}([1, 0, 0]^T, [0, 1, 0]^T, [0, 0, 1]^T) \). In Fig. 2, the probability functions with respect to which the components \( p_i \) are conservative are enclosed by the egg shape curves. In Fig. 3, the probability functions which are conservative with respect to the components \( p_i \) lie in the half-plane that does not contain the uniform function \( p_u \), \( p_u = \left[ \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right] \), which is universally conservative.
First, note that $p_2$ is conservative with respect to $p_1$ with zero conservativeness $C(p_1||p_2)$. In fact, $p_2$ has the uniform distribution for two states only, and thus, restricting to the case $P_1 = 0$, it is universally conservative. Next, $p_4$ is conservative with respect to $p_1$ with nonzero, i.e., positive, conservativeness $C(p_1||p_4)$. Further, $p_3$ is conservative with respect to $p_2$, $C(p_2||p_3) > 0$. Finally, no other pairs are conservative.

The lack of transitivity is documented by the densities $p_1$, $p_2$ and $p_3$, since $C(p_1||p_2) = 0$, $C(p_2||p_3) > 0$, but $C(p_1||p_3) < 0$. The lack is comprehensible, making an acceptably crude approximation of an acceptably crude approximation needs not result in an acceptable approximation.

Note also that $p_4 \in M(p_1, p_5)$, i.e., $p_1$, $p_4$ and $p_5$ are linearly dependent. Regarding the components $p_2$ and $p_5$, the only probability function that is conservative with respect to both components is the universally conservative function $p_u$. The component $p_1$ is conservative only with respect to itself, since there is no uncertainty in $p_1$.

The densities that are conservative with respect to all selected components are given by the intersection of the corresponding egg shapes in Fig. 2. The intersections are convex sets with two vertices, the first one is given by the universally conservative probability function $p_u$, the second one is given by the maximum entropy mixture $\pi^*$. Remark that the maximum entropy mixture is also the minimum entropy conservative density, because, rather simplistically said, the entropy decreases with the distance from the centre at $p_u$.

If one component is conservative with respect to the second one in a two component case, the maximum entropy mixture is given by the conservative component, $\pi^*_{1,2} = p_2$, $\pi^*_{1,4} = p_4$, $\pi^*_{2,3} = p_3$. For linearly dependent components, the maximum entropy mixture can be expressed by different combinations, $\pi^*_{1,4,5} = \pi^*_{1,5} = \pi^*_{4,5}$, and the conservativeness with respect to the components is zero, especially note that $C(p_1||\pi^*_{1,4,5}) = 0$.

In the three state problem, the pairwise maximum entropy mixture are the maximum entropy mixture for more components, for example $\pi^*_{1,2,3} = \pi^*_{1,3} = \pi^*_{2,3} = \pi^*_{4,5}$. The conservativeness with respect to the redundant components is positive, $C(p_2||\pi^*_{1,3}) > 0$, $C(p_2||\pi^*_{1,4}) > 0$, $C(p_2||\pi^*_{4,5}) > 0$, and therefore these components can be pruned.

From Fig. 3, it can be seen that the straight lines that divides the probability simplex to the conservative and nonconservative half-planes rotate with the weight of a mixture. For example, the linear classifier at $p_3$ is horizontal and rotates anti-clockwise to $\pi^*_{1,5}$ and further to $p_4$ where it is nearly vertical. It is easy to show by contradiction that the classifiers are tangential to the entropy contours at the level of the current entropy. If they were secant, there would be higher entropy densities with respect to which a density would be conservative.

C. Overlapping Gaussian hypotheses

The last example inspects two Gaussian densities $p_1 = N(\mu_1, \Sigma_1)$, $i = 1, 2$, that overlap significantly. Let the densities be $p_1 = N(0, 1)$ and $p_2 = N(3, 4)$.

Fig 4 shows the conservativeness (3) of the mixtures $\pi = \omega p_1 + (1-\omega)p_2$ with respect to the components. The integrals in (4), (5) are solved numerically. The values for $\omega = 0$ and $\omega = 1$ can be determined analytically, because the divergence and the entropy are known for Gaussians. The conservativeness of a density with respect to itself is zero, $C(p_2||\pi^{\omega = 0}) = 0$, $C(p_1||\pi^{\omega = 1}) = 0$. The conservativeness of two Gaussians is given by $C(p_2||\pi) = \frac{1}{2} \text{tr}(\{\Sigma_2 - \Sigma_1 - (\mu_2 - \mu_1)^T (\mu_2 - \mu_1)^T\} P^{-1})$, where tr denotes the trace of the matrix. The resemblance to (1) is not accidental. The equivalence of the sign for one dimensional Gaussians has been the basic requirement [12]. Therefore it holds $C(p_1||\pi^{\omega = 0}) = -1.625$ and $C(p_2||\pi^{\omega = 1}) = -9.5$. Remark also that $\mathcal{H}(\pi^{\omega = 0}) = \mathcal{H}(p_1) \approx 1.42$ and $\mathcal{H}(\pi^{\omega = 1}) = \mathcal{H}(p_2) \approx 2.11$.

According to the theoretical results of this paper, the mixture $\pi^*$ which maximise the entropy $\mathcal{H}(\pi)$ is conservative with respect to both components with zero conservativeness. The corresponding weight $\omega^*$ is approximately equal to 0.3. Since the densities overlap, the approximation (23) is crude. The approximate weight $\omega^*$ is given by $\omega^* = \frac{1}{3}$.

Fig 5 compares the components with the merged densities. As it has been shown in Section VI-A, the weighted functions $\omega^* p_1$, $(1 - \omega^*) p_2$, have the same maxima. But because the components $p_1$, $p_2$ overlap significantly, the resulting mixture
is not conservative with respect to both components. It is conservative with respect to the first density, $C(p_1 || \pi^a) = 0.071$, but not to the second one, $C(p_2 || \pi^a) = -0.036$. Note that the approximate mixture $\pi^*$ is biased towards the first component, $\omega^a > \omega^b$.

The weighted functions $\omega^a p_1, (1 - \omega^a)p_2$ of the maximum entropy mixture $\pi^*$ do not have the same maxima, which they have in the case the Gaussians are distant. But the maximum entropy mixture does tend towards a uniform density over the areas of nonnegligible values of probability densities.

VII. SUMMARY

Using the recently proposed definition of a conservative density, a density which is conservative with respect to all considered unweighted hypotheses has been searched. The key finding is that such a density has to be conservative with respect to all mixtures of the hypotheses, i.e. its entropy cannot be lower than the entropy of an arbitrary mixture. It has been shown that the mixture with the maximum entropy is conservative with respect to all hypotheses and therefore it is the solution with the lowest entropy. The conservativeness of the maximum entropy mixture with respect to the components that can have positive weights is zero. If the supports of the hypotheses are disjoint, the weights of the maximum entropy mixture are proportional to the exponential function of the entropy, i.e. to the effective support volumes. Generally, the conservative merging provides a density that tends towards a uniform density which is universally conservative. Moreover, considering the piecewise constant density analogy, the approach is also applicable to discrete random variables.

ACKNOWLEDGMENT

This work was supported by the Czech Science Foundation, project no. P103/11/1353 and by the European Regional Development Fund (ERDF), project “NTIS – New Technologies for Information Society”, European Centre of Excellence, CZ.1.05/1.1.00/02.0090.

REFERENCES