Covariance Intersection in Nonlinear Estimation based on Pseudo Gaussian Densities

Benjamin Noack, Marcus Baum, and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
Institute for Anthropomatics,
Karlsruhe Institute of Technology (KIT), Germany.
Email: benjamin.noack@ieee.org, marcus.baum@kit.edu, uwe.hanebeck@ieee.org

***Abstract*—Many modern fusion architectures are designed to process and fuse data in networked systems. Alongside the advantages, such as scalability and robustness, distributed fusion techniques particularly have to tackle the problem of dependencies between locally processed data. In linear estimation problems, uncertain quantities with unknown cross-correlations can be fused by means of the covariance intersection algorithm, which avoids overconfident fusion results. However, for nonlinear system dynamics and sensor models perturbed by arbitrary noise, it is not only a problem to characterize and parameterize dependencies between estimates, but also to find a proper notion of consistency. This paper addresses these issues by transforming the state estimates to a different state space, where the corresponding densities are Gaussian and only linear dependencies between estimates, i.e., correlations, can arise. These pseudo Gaussian densities then allow the notion of covariance consistency to be used in distributed nonlinear state estimation.**

Keywords: Covariance intersection, covariance bounds, distributed estimation, decentralized estimation, nonlinear estimation, state transformation.

I. Introduction

Bayesian state estimation techniques provide the means to derive a probabilistic description of an uncertain state from noisy measurements and input signals. The rapid advances in sensor and communication technology entail an increasing demand for implementing these estimation algorithms in distributed networked systems [1]–[3]. The general idea is that data is collected and processed locally on different sensor nodes, with the aim of monitoring large-scale phenomena, distributing computational resources, and increasing robustness to failures. In order to form a global estimate, each node communicates its locally processed data to other nodes, where the local estimates can then be fused. For the purpose of scalability, the nodes are generally intended to operate independently, but this does not imply that the processed data are independent of each other. The reasons for dependencies between the local data sets are manifold: The nodes may share common prior information, local state predictions may be affected by the same process noise, sensor noises may be correlated, etc. All told, the central problem is that local processing and communication may lead to "data incest" [4], i.e., double counting of information. Apparently, the best fusion results can only be achieved, if the common information between two nodes is stored somewhere and can be removed from the fused estimate, but this imposes additional requirements on the network topology and may limit its flexibility and scalability significantly.

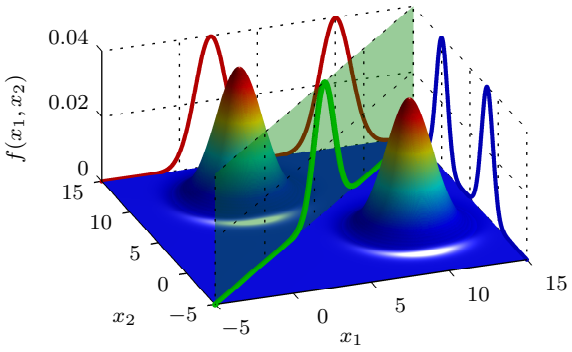


Figure 1. The dependency structure between the two local bimodal estimates $f(x_1)$ (red) and $f(x_2)$ (blue) is characterized by the joint density $f(x_1, x_2)$. The fusion result is the normalized green density that lies on the diagonal of the joint density.

In many existing fusion architectures, hierarchical topologies are employed, which allow for explicitly exploiting dependencies between estimates [1]–[3], [5]. However, in general, it is preferable that the nodes do not need to be aware of the network topology. In such a fully distributed/decentralized network, whose topology can be arbitrary and can change over time [6], only suboptimal estimates can then be computed. For linear estimation problems, the covariance intersection (CI) algorithm [7]–[9] has been proposed, which yields fusion results that do not underestimate the actual mean-squared-error (MSE) matrix. While CI considers a convex combination of the estimates in the information space, it can also be formulated as a bound on the covariance matrix in the joint state space [10], [11]. More precisely, both fusion techniques provide a parametric family of fusion results. The corresponding weighting parameter is commonly determined to minimize the determinant or trace of the fused MSE matrix. Other optimization criteria have been analyzed against the background of information theory [12], [13], set theory [6], [14], or approximate solutions [4]. When the estimates to be fused share a common prior, also tighter bounds can be found [15]. Eventually, CI has proven to be a powerful tool in decentralized fusion architectures provided that only linear system and sensor models are considered.

In nonlinear decentralized estimation problems, it is still an open question what is to be considered as a conservative fusion result, when dependencies are unknown. Fig. 1 shows a possible dependency structure of two bimodal densities. Whereas in linear estimation stochastic dependencies can be entirely parameterized by their cross-correlations, dependencies between estimates in nonlinear estimation can be arbitrary. Even for local Gaussian estimates, the dependencies are not necessarily linear, when the system and sensor models are nonlinear. Beginning from information-theoretic considerations in [12], CI has been generalized to exponential mixture densities (EMD) [16]–[18]. Although EMDs provide promising results, it remains questionable whether EMDs can tackle the arbitrariness of possible dependencies. In this paper, we will analyze this issue by transforming the state space into a different space, where the models become linear and only linear dependencies can arise. There, we can parameterize dependencies, can apply the CI algorithm, and can employ the same notion of consistency as in linear estimation problems. We will study the relation to EMDs and will evaluate the proposed concept in simulations.

II. Preliminaries

In this paper, underlined variables $\underline{x}$ denote vectors and lowercase boldface letters $\underline{\boldsymbol{x}}$ are used for random quantities. Matrices are written in uppercase boldface letters $\mathbf{C}$. By $(\underline{\hat{x}}, \mathbf{C})$, we denote an estimate with mean $\underline{\hat{x}}$ and covariance matrix $\mathbf{C}$. The function $\mathcal{N}(\cdot; \underline{\hat{x}}, \mathbf{C})$ is the Gaussian density

$$f(\cdot) = c \cdot \exp\left\{-\tfrac{1}{2}\big(\cdot - \underline{\hat{x}}\big)^{\mathrm{T}} (\mathbf{C})^{-1} \big(\cdot - \underline{\hat{x}}\big)\right\} \ .$$

with mean $\underline{\hat{x}}$ and covariance matrix $\mathbf{C}$. c is the normalization factor. By

$$\tilde{\mathbf{C}} \geq \mathbf{C} \text{ or } \tilde{\mathbf{C}} - \mathbf{C} \geq \mathbf{0} \ ,$$

we denote that $\tilde{\mathbf{C}} - \mathbf{C}$ is a positive semidefinite matrix.

III. Problem Statement

In a sensor network, let A and B denote two fusion nodes that communicate their estimates of an uncertain system state $\underline{\boldsymbol{x}}_k$ at a given time instant k. To simplify matters, we omit the time index in the following considerations. The sets $\mathcal{Z}^A$ and $\mathcal{Z}^B$ are the information sets collected at each local node. They consists of the measurement sequences obtained through own sensor observations and through communication with other nodes. The local state estimates are represented by the conditional densities $f(\underline{x} \,|\, \mathcal{Z}^A)$ and $f(\underline{x} \,|\, \mathcal{Z}^B)$, which can be fused according to

$$\begin{aligned} f(\underline{x} \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) &= \frac{f(\mathcal{Z}^A \cup \mathcal{Z}^B | \underline{x}) \cdot f(\underline{x})}{f(\mathcal{Z}^A \cup \mathcal{Z}^B)} \\ &= \frac{f(\mathcal{Z}^A \,|\, \underline{x}) \cdot f(\mathcal{Z}^B \,|\, \underline{x}) \cdot f(\underline{x})}{f(\mathcal{Z}^A \cup \mathcal{Z}^B)} \\ &= \frac{f(\mathcal{Z}^A) \cdot f(\mathcal{Z}^B)}{f(\mathcal{Z}^A \cup \mathcal{Z}^B)} \cdot \frac{f(\underline{x} \,|\, \mathcal{Z}^A) \cdot f(\underline{x} \,|\, \mathcal{Z}^B)}{f(\underline{x})} \ , \end{aligned} \tag{1}$$

if $\mathcal{Z}^A$ and $\mathcal{Z}^B$ are conditionally independent [2], [5], i.e., $f(\mathcal{Z}_k^A \cup \mathcal{Z}^B \,|\, \underline{x}) = f(\mathcal{Z}^A \,|\, \underline{x}) \cdot f(\mathcal{Z}^B \,|\, \underline{x})$. When no prior information $f(\underline{x})$ on $\underline{\boldsymbol{x}}$ is available, the fusion formula reduces to

$$f(\underline{x} \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) = c \cdot f(\underline{x} \,|\, \mathcal{Z}^A) \cdot f(\underline{x} \,|\, \mathcal{Z}^B) \ ,$$

where c is the normalization factor. The conditional densities $f(\underline{x} \,|\, \mathcal{Z}^A)$ and $f(\underline{x} \,|\, \mathcal{Z}^B)$ can be considered as the marginals of the joint density[1]

$$f(\underline{x}_1, \underline{x}_2 \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) := f(\underline{x}_1 \,|\, \mathcal{Z}^A) \cdot f(\underline{x}_2 \,|\, \mathcal{Z}^B) \ . \tag{2}$$

The joint density is related to the fusion result (1) by

$$\begin{aligned} f(\underline{x} \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) &= c \cdot f(\underline{x} \,|\, \mathcal{Z}^A) \cdot f(\underline{x} \,|\, \mathcal{Z}^B) \\ &= c \cdot f(\underline{x}, \underline{x} \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) \ . \end{aligned} \tag{3}$$

Thus, the fused density (1) is obtained from the joint density conditioned on the event $\mathcal{E} = \{[\underline{x}_1, \underline{x}_2] \mid \underline{0} = \underline{x}_1 - \underline{x}_2\}$ that $\underline{x}_1$ and $\underline{x}_2$ are equal.

In general, the information sets of two sensor nodes are not conditionally independent of each other. In analogy to (1), the fused estimated probability density is then given[2] by

$$\begin{aligned} f(\underline{x} | \mathcal{Z}^A \cup \mathcal{Z}^B) &= \frac{f(\mathcal{Z}^A \cup \mathcal{Z}^B \backslash \mathcal{Z}^A | \underline{x}) \cdot f(\underline{x})}{f(\mathcal{Z}^A \cup \mathcal{Z}^B)} \\ &= \frac{f(\mathcal{Z}^A) \cdot f(\mathcal{Z}^B \backslash \mathcal{Z}^A)}{f(\mathcal{Z}^A \cup \mathcal{Z}^B)} \cdot \frac{f(\underline{x} | \mathcal{Z}^A) \cdot f(\underline{x} | \mathcal{Z}^B \backslash \mathcal{Z}^A)}{f(\underline{x})} \end{aligned}$$

and, according to (2), the joint density is

$$f(\underline{x}_1, \underline{x}_2 \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) := f(\underline{x}_1 \,|\, \mathcal{Z}^A) \cdot f(\underline{x}_2 \,|\, \mathcal{Z}^B \backslash \mathcal{Z}^A) \ . \tag{4}$$

Again the fusion result lies on the *diagonal*

$$\begin{aligned} f(\underline{x} \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) &= c \cdot f(\underline{x} \,|\, \mathcal{Z}^A) \cdot f(\underline{x} \,|\, \mathcal{Z}^B \backslash \mathcal{Z}^A) \\ &= c \cdot f(\underline{x}, \underline{x} \,|\, \mathcal{Z}^A \cup \mathcal{Z}^B) \end{aligned} \tag{5}$$

of the joint density. An example is depicted in Fig. 1, where the fusion of two bimodal and even equal densities $f(\underline{x} \,|\, \mathcal{Z}^A)$ and $f(\underline{x} \,|\, \mathcal{Z}^B)$ unexpectedly yields a unimodal density. The fusion result is plotted in Fig. 2. Hence, the local information

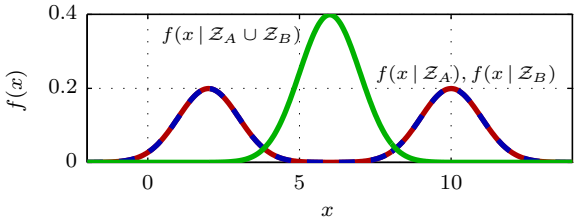


Figure 2. The fusion of the blue and red density $f(x|\mathcal{Z}_A)$ and $f(x|\mathcal{Z}_B)$ (red/blue dashed) yields the unexpected green density.

may differ severely from the global estimate. Especially in a fully decentralized network, only the marginals $f(\underline{x} \,|\, \mathcal{Z}^A)$ and $f(\underline{x} \,|\, \mathcal{Z}^B)$ are known, i.e., the dependency structure between

[1] The joint density can also directly be derived from (1), when no prior densities $f(\underline{x}_1)$, $f(\underline{x}_2)$, and $f(\underline{x}_1, \underline{x}_2)$ are given.

[2] Essentially, the conditional independence of $\mathcal{Z}^A$ and $\mathcal{Z}^B \backslash \mathcal{Z}^A$ only holds for static or deterministic systems. For dynamic stochastic systems, we can draw the same conclusions by replacing $\mathcal{Z}^B \backslash \mathcal{Z}^A$ by an artificially introduced set $\tilde{\mathcal{Z}}^B$ that encompasses the conditionally independent information.

local estimates remains hidden and the joint density or respectively the conditional density $f(\underline{x}|\mathcal{Z}^B\backslash\mathcal{Z}^A)$ therefore cannot be reconstructed uniquely. In other words, finding the joint density for given marginals is an ill-posed inverse problem. Even for local Gaussian estimates, the fusion result can be far from being Gaussian anymore [19].

As stated in the introduction, the reasons for the lack of independence are manifold. The good news is that the dependencies are not as arbitrary as they might appear at first sight: Our knowledge about the sensor, system, and fusion models limits the arbitrariness *and* we know how to systematically deal with unknown correlations in linear estimation problems.

IV. State of the Art

The problem of fusing two estimates has been investigated intensely in linear contexts. For linear system dynamics and sensor models, where all participating noise terms are assumed to be Gaussian, only linear dependencies between estimates can arise and need to be considered. The linearity of every processing step implies that the joint density (4) is also Gaussian and therfore can be parameterized by its mean vector $\hat{\underline{x}}$ and covariance matrix $\mathbf{C}$. At a given and fixed time instant k, let $f(\underline{x}\,|\,\mathcal{Z}^A) = \mathcal{N}(\underline{x};\hat{\underline{x}}_A,\mathbf{C}_A)$ and $f(\underline{x}\,|\,\mathcal{Z}^B) = \mathcal{N}(\underline{x};\hat{\underline{x}}_B,\mathbf{C}_B)$ be the locally estimated densities. If the fusion nodes are aware of the cross-covariance matrix $\mathbf{C}_{AB}$, the estimates can then be fused according to the Bar-Shalom/Campo combination rule [20]. Its result is to be regarded as the best linear unbiased estimate (BLUE), when no prior information on $\underline{\boldsymbol{x}}$ is available and can be incorporated. In the situation when the cross-covariance matrix is unknown, the covariance intersection (CI) algorithm [7]–[9] enables us to compute a *conservative* estimate with mean

$$\hat{\underline{x}}_{\mathrm{CI}} = \mathbf{C}_{\mathrm{CI}}\big(\omega\mathbf{C}_A^{-1}\hat{\underline{x}}_A + (1-\omega)\mathbf{C}_B^{-1}\hat{\underline{x}}_B\big) \tag{6}$$

and covariance matrix

$$\mathbf{C}_{\mathrm{CI}} = \big(\omega\mathbf{C}_A^{-1} + (1-\omega)\mathbf{C}_B^{-1}\big)^{-1}\ , \tag{7}$$

where $\omega\in[0,1]$ is a weighting factor. Here, an estimate $(\hat{\underline{x}},\mathbf{C})$ is called conservative, if it preserves covariance consistency

$$\mathbf{C} - \widetilde{\mathbf{C}} \geq 0\ ,$$

where $\widetilde{\mathbf{C}}$ denotes the actual MSE matrix $\mathrm{E}[\underline{e}_x\underline{e}_x^{\mathrm{T}}]$ with $\underline{e}_x = \underline{x}-\hat{\underline{x}}$. The CI algorithm yields covariance consistent estimates for every $\omega\in[0,1]$ and every possible cross-covariance matrix $\mathbf{C}_{AB}$ provided that $(\hat{\underline{x}}_A,\mathbf{C}_A)$ and $(\hat{\underline{x}}_B,\mathbf{C}_B)$ are consistent estimates. The CI algorithm can alternatively be formulated in the joint state space: The covariance bounds (CB) algorithm [10], [11] computes an upper bound for the joint covariance matrix, i.e.,

$$\begin{bmatrix} \frac{1}{\omega}\cdot\mathbf{C}_A & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\omega)}\cdot\mathbf{C}_B \end{bmatrix} \geq \begin{bmatrix} \mathbf{C}_A & \mathbf{C}_{AB} \\ \mathbf{C}_{BA} & \mathbf{C}_B \end{bmatrix} \tag{8}$$

with $\omega\in[0,1]$. The corresponding joint density is then given by

$$\mathcal{N}\left(\begin{bmatrix}\underline{x}_1\\ \underline{x}_2\end{bmatrix};\begin{bmatrix}\hat{\underline{x}}_A\\ \hat{\underline{x}}_B\end{bmatrix},\begin{bmatrix} \frac{1}{\omega}\cdot\mathbf{C}_A & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\omega)}\cdot\mathbf{C}_B \end{bmatrix}\right)$$

and, in compliance with (3), the fusion rule

$$f(\underline{x}\,|\,\mathcal{Z}_A\cup\mathcal{Z}_B) = c\cdot\mathcal{N}(\underline{x};\hat{\underline{x}}_A,\tfrac{1}{\omega}\mathbf{C}_A)\cdot\mathcal{N}(\underline{x};\hat{\underline{x}}_B,\tfrac{1}{(1-\omega)}\mathbf{C}_B)$$

can be applied. This approach yields the same estimated mean (6) and covariance matrix (7) as CI. Besides many other optimization criteria, the weight ω is generally calculated to minimize $\det(\mathbf{C}_{\mathrm{CI}})$, which also corresponds to the covariance ellipsoid with minimum volume. Since CI can result into highly conservative estimates, it has been developed further in order to exploit additional knowledge on the correlation structure. One way consists in decomposing the estimates into dependent and independent parts. Then, CI only needs to be applied to the dependent parts, which is called split covariance intersection [9], [21]. Alternatively, symmetric or asymmetric bounds on $\mathbf{C}_{AB}$ in (8) can be employed [11], [22].

Like the Kalman filter, CI only relies on the assumption of linearity and the first two moments of $f(\underline{x}\,|\,\mathcal{Z}^A)$ and $f(\underline{x}\,|\,\mathcal{Z}^B)$ [18]. However, in many nonlinear estimation problems, linear error statistics are no longer sufficient when for instance, the multimodality of the estimated density is to be maintained. Mahler [23] and Hurley [12] have independently proposed to generalize CI to the exponential mixture density (EMD)

$$f_\omega(\underline{x}|\mathcal{Z}_A\cup\mathcal{Z}_B) = \frac{f^\omega(\underline{x}|\mathcal{Z}_A)\cdot f^{(1-\omega)}(\underline{x}|\mathcal{Z}_B)}{\int_{\mathbb{R}^N} f^\omega(\underline{x}|\mathcal{Z}_A)\cdot f^{(1-\omega)}(\underline{x}|\mathcal{Z}_B)\,\mathrm{d}\underline{x}} \tag{9}$$

with $\omega\in[0,1]$. Julier [18] has shown that this fusion formula counteracts the double counting of information and discussed in [16], in which way EMDs are to be considered as conservative and consistent estimates. Promising results have been presented with regard to Gaussian mixtures [17], exponentials of polynomials [24], and multi-object densities [25]. Nevertheless, it still remains an open question what is to be considered as a conservative estimate in nonlinear estimation. For example, applying the EMD formula (9) to fuse the blue and red density in Fig. 2 yields again the same bimodal density, which hardly captures the mode of the green true fusion result. In the following, we address this problem by transforming the state to a space, where the models become linear or almost linear and therefore, we can apply the notion of covariance consistency. In the transformed state space, dependencies cannot be arbitrary, but must be linear.

V. Nonlinear Decentralized State Estimation based on Pseudo Gaussian Densities

The general idea behind this section is to regard the true estimated density as a (pseudo) Gaussian density in a different state space [26]. Against the background of set-theoretic state estimation, this approach has been employed to represent complicated sets as ellipsoids [27] and, in [28], this technique is referred to as non-minimal state Kalman filtering. In particular, this concept is strongly related to Carleman (bi-)linearizations [29]–[31].

The underlying mechanism relies on a proper transformation $\underline{T}:\mathcal{S}\to\mathcal{S}^*$ with $\mathcal{S}\subseteq\mathbb{R}^N$, $\mathcal{S}^*\subseteq\mathbb{R}^M$, and

$$\underline{x}^* = \underline{T}(\underline{x}) = [T_1(\underline{x}),\ldots,T_M(\underline{x})]^{\mathrm{T}}\ , \tag{10}$$

so that the state estimates become normally distributed in $\mathcal{S}^*$, i.e., $\mathcal{N}(\underline{x}^*; \hat{\underline{x}}^*, \mathbf{C}^*)$. The fundamental prerequisite is that the fusion of two estimates and the measurement update turn into linear operations. The former requirement is ensured, if the dependencies between the estimated pseudo Gaussian densities are solely linear. The latter condition is met, if a nonlinear sensor model

$$\hat{\underline{z}}_k = h(\underline{\boldsymbol{x}}_k) + \underline{\boldsymbol{v}}_k \tag{11}$$

can be rewritten as a linear update equation

$$\hat{\underline{z}}^*_k = \mathbf{H}^* \, \underline{\boldsymbol{x}}^*_k + \underline{\boldsymbol{v}}^*_k \tag{12}$$

in the transformed state space $\mathcal{S}^*$, where $\underline{\boldsymbol{v}}_k$ is an additive noise term, which becomes a normally distributed perturbation $\underline{\boldsymbol{v}}^*_k$.

For static systems, where observations are only perturbed by additive Gaussian noise, a proper transformation $\underline{T}$ can easily be determined. Let h^i with $i \in \{1, \ldots, L\}$ denote the L different sensor models in a given network. Then, the estimation process becomes linear by employing the transformed state

$$\underline{x}^* = T(\underline{x}) := [h^1(\underline{x}), \ldots, h^L(\underline{x})]^{\mathrm{T}} \ , \tag{13}$$

where a single node with sensor model $h^i(\underline{x})$ partially observes $\underline{x}^*$, i.e., the ith component of $\underline{x}^*$. Hence, communication can only cause linear dependencies between local estimates. In general, the dimension of the transformed state space is higher than the dimension of the original space and, of course, the choice of $\underline{T}$ is not unique, since $\underline{x}^*$ can, for instance, be expanded by any linear combination of partial states $\underline{x}^*_i$ and $\underline{x}^*_j$. In a situation where combinations of the state $\underline{x}$ remain unobserved, the transformed state space can be lower-dimensional. For example, when a two-dimensional state is observed by a single distance sensor, a transformation to a one-dimensional state is sufficient.

When dealing with dynamic systems, we aspire to find a mapping $\underline{T}$ that also transforms a nonlinear system model

$$\underline{\boldsymbol{x}}_{k+1} = a(\underline{\boldsymbol{x}}_k, \hat{\underline{u}}_k, \underline{\boldsymbol{w}}_k) \tag{14}$$

to a linear state evolution

$$\underline{\boldsymbol{x}}^*_{k+1} = \mathbf{A}^* \, \underline{\boldsymbol{x}}^*_k + \mathbf{B}^*_k \, (\hat{\underline{u}}^*_k + \underline{\boldsymbol{w}}^*_k) \ ,$$

where $\hat{\underline{u}}_k$ and $\hat{\underline{u}}^*_k$ comprise possible input quantities in the original and the transformed state space, respectively. The system noise $\underline{\boldsymbol{w}}_k$ becomes a normally distributed disturbance term $\underline{\boldsymbol{w}}^*_k$. On the assumption that a mapping $\underline{T}$ exists that simultaneously linearizes the system and sensor models, exclusively linear dependencies between the local estimates can arise. Hence, CI can be employed for suboptimal decentralized estimation based upon pseudo Gaussian densities, which is considered more closely in Subsection V-A. Unfortunately, such a transformation $\underline{T}$ to a finite-dimensional space, where both the system dynamics and the sensor models are linear, cannot be constructed in general. Therefore, linear approximations of systems through state transformations are discussed in Subsection V-B. All told, even if no optimal transformation to a linear form of the system model is obtained, the measurement update and fusion are now linear.

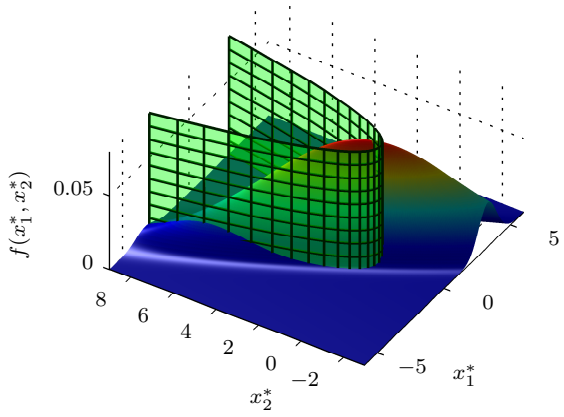


Figure 3. Estimated pseudo Gaussian density in sensor node 1. CI has been used to fuse estimates. The true density lies on the green manifold.

The original nonlinearities come into play whenever an estimate in the original state space $\mathcal{S}$ has to be computed. Of course, the reverse transformation may involve complicated nonlinear estimation techniques, but the required calculations can be performed independently from the state estimation process, which solely relies on the pseudo Gaussian densities and the transformed models. For a sensor network, this further implies that every node can process the pseudo Gaussian estimates locally and linearly. In the data sink, where in general higher computational power is available, the true density and the desired parameters can be computed from the pseudo Gaussian estimates. The original density $f(\underline{x}_k)$ is related to the estimate $(\hat{\underline{x}}^*_k, \mathbf{C}^*_k)$ by

$$\begin{aligned} f(\underline{x}_k) &= c \cdot \mathcal{N}(T(\underline{x}_k), \hat{\underline{x}}^*, \mathbf{C}^*) \\ &= \tilde{c} \cdot \exp\left\{ \tfrac{1}{2} \big(T(\underline{x}_k) - \hat{\underline{x}}^*_k\big)^{\mathrm{T}} (\mathbf{C}^*_k)^{-1} \big(T(\underline{x}_k) - \hat{\underline{x}}^*_k\big) \right\} \ . \end{aligned} \tag{15}$$

Note that this calculation can be regarded as a measurement update with $\hat{\underline{x}}^*_k = T(\underline{x})$, where no prior is available.

A. Covariance Intersection for Pseudo Gaussian Densities

For the purpose of fusing two local estimates $f(\underline{x}|\mathcal{Z}_A)$ and $f(\underline{x}|\mathcal{Z}_B)$, where $\mathcal{N}(T(\underline{x}); \hat{\underline{x}}^*_A, \mathbf{C}^*_A)$ and $\mathcal{N}(T(\underline{x}); \hat{\underline{x}}^*_B, \mathbf{C}^*_B)$ are the corresponding pseudo Gaussian densities, we can now apply the CI equations (6) and (7). By virtue of the preceding considerations, the dependencies are linear and CI provides a covariance consistent suboptimal fusion result $(\hat{\underline{x}}^*_{\mathrm{CI}}, \mathbf{C}^*_{\mathrm{CI}})$ in the transformed state space. The following example gives an impression of the presented concept.

Example: Simple Sensor Network

We consider a one-dimensional setup with three closely positioned sensor nodes at $P^1 = -1$, $P^2 = -1.2$, and $P^3 = -0.5$. They measure the distances

$$\hat{z}^i_k = (\boldsymbol{x} - P^i)^2 + \boldsymbol{v}^i \ , \ i \in \{1, 2, 3\} \ ,$$

where $\boldsymbol{v}^i$ are zero-mean Gaussian noise terms with high variances 5, 9, and 7. The true state is located at 1. The prior

estimate at every node is $\hat{x}_0 = -2$ with variance $\mathbf{C}_0 = 5$. The measurement equation can be transformed to the affine model

$$\hat{z}_k^i = \mathbf{H}^{*,i} \cdot \underline{\boldsymbol{x}}_k^* + (P^i)^2 + \boldsymbol{v}^i$$

with the linear mapping $\mathbf{H}^{*,i} = [-2\,P^i, 1]$ and the transformed state $\underline{\boldsymbol{x}}_k^* = [\boldsymbol{x}_1^*, \boldsymbol{x}_2^*]^{\mathrm{T}} = [\boldsymbol{x}_k, \boldsymbol{x}_k^2]^{\mathrm{T}}$. Fig. 3 shows the bivariate Gaussian estimate $(\underline{\hat{x}}_5^{*,1}, \mathbf{C}_5^{*,1})$ of sensor 1 after 5 measurement steps, where the sensors have communicated their results at every time step by means of the CI algorithm. The density of the estimate in the original state space lies on the manifold, which corresponds to the green density in Fig. 4. There, different fusion results are compared. The blue density represents the optimal Bayesian fusion result, where all dependencies are known. It captures the true state well, in contrast to the red density, where dependencies have been ignored. Evidently, the CI algorithm still provides a good fusion result.

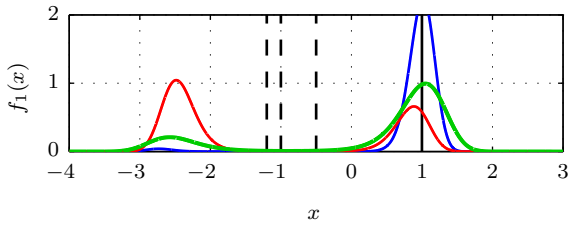


Figure 4. The density that corresponds to the pseudo Gaussian density in Fig. 3 is drawn in green. The blue density belongs to the pseudo Gaussian estimate, where correlations are known and the Bar-Shalom/Campo combination rule has been applied. It is the optimal Bayesian estimate. For the red density, the naïve fusion rule has been used. The dashed black lines are the sensor positions and the solid black line shows the true state.

The example elucidates how CI applied in the transformed space $\mathcal{S}^*$ flattens the modes of the density in the original state space $\mathcal{S}$. Throughout this paper, the parameter ω for the fusion formula (7) is chosen to minimize $\det(\mathbf{C}_{\mathrm{CI}}^*)$. In terms of the underlying densities, the CI algorithm in $\mathcal{S}^*$ can be expressed as

$$\begin{aligned} &f_\omega(\underline{x} \,|\, \mathcal{Z}_A \cup \mathcal{Z}_B) \\ &= c \cdot \mathcal{N}(T(\underline{x}); \underline{\hat{x}}_A^*, \tfrac{1}{\omega}\mathbf{C}_A^*) \cdot \mathcal{N}(T(\underline{x}); \underline{\hat{x}}_B^*, \tfrac{1}{(1-\omega)}\mathbf{C}_B^*) \\ &= c \cdot (\mathcal{N}(T(\underline{x}); \underline{\hat{x}}_A^*, \mathbf{C}_A^*))^{\omega} \cdot (\mathcal{N}(T(\underline{x}); \underline{\hat{x}}_B^*, \mathbf{C}_B^*))^{(1-\omega)} \quad , \end{aligned}$$

which is the EMD update rule (9) directly applied to the densities $f(\underline{x} \,|\, \mathcal{Z}_A)$ and $f(\underline{x} \,|\, \mathcal{Z}_B)$ to be fused. In this regard, EMDs actually provide consistent estimates in the sense of covariance consistency in $\mathcal{S}^*$, which confirms the considerations in [16]. Maybe, this also points to the direction to define consistency in nonlinear estimation problems by means of state transformations, which linearize the estimation process. Another valuable feature of the space $\mathcal{S}^*$ is that it enables us to parameterize every possible dependency between two estimates $(\underline{\hat{x}}_A^*, \mathbf{C}_A^*)$ and $(\underline{\hat{x}}_B^*, \mathbf{C}_B^*)$ by the cross-covariance matrix $\mathbf{C}_{AB}^*$. Possible dependencies are illustrated in the following example, which shall conclude this section.

Example: Possible Fusion Results

As in the above example, we consider the transformation $\underline{\boldsymbol{x}}_k^* = [\boldsymbol{x}_1^*, \boldsymbol{x}_2^*]^{\mathrm{T}} = [\boldsymbol{x}_k, \boldsymbol{x}_k^2]^{\mathrm{T}}$. In $\mathcal{S}^*$, the local estimates are

$$(\underline{\hat{x}}_A^*, \mathbf{C}_A^*) = \left(\begin{bmatrix} 4 \\ 4 \end{bmatrix}, 3 \cdot \begin{bmatrix} 7 & 1 \\ 1 & 3 \end{bmatrix} \right) \tag{16}$$

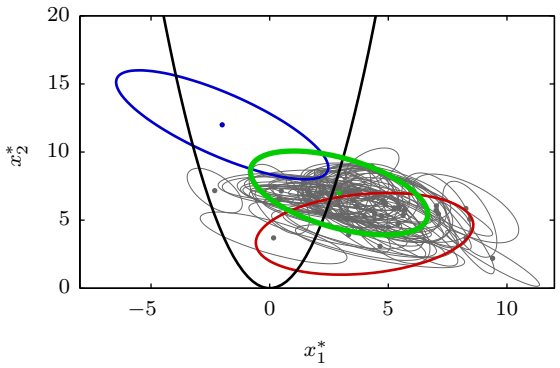


Figure 5. Covariance ellipsoids of the estimates in $\mathcal{S}^*$. The red and the blue estimate are fused according to the Bar-Shalom/Campo rule for different possible correlations (gray thin ellipsoids). The green ellipsoid corresponds to the CI algorithm. The densities in Fig. 6 lie on the black parabola.

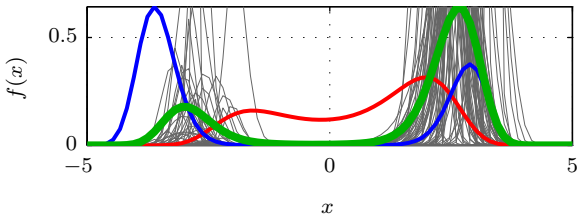


Figure 6. Probability density functions in the original space $\mathcal{S}$.

and

$$(\underline{\hat{x}}_B^*, \mathbf{C}_B^*) = \left(\begin{bmatrix} -2 \\ 12 \end{bmatrix}, 2 \cdot \begin{bmatrix} 10 & -7 \\ -7 & 8 \end{bmatrix} \right) , \tag{17}$$

which are plotted in Fig. 5 as the red and the blue covariance ellipsoid, respectively. 400 possible cross-covariance matrices $\mathbf{C}_{AB}^*$ have been determined randomly. For these, the Bar-Shalom/Campo rule has been applied to fuse (16) and (17). In Fig. 5 and 6, the according results are drawn in gray. With CI, the green ellipsoid has been computed, which correspond to the green EMD in Fig. 6. As can be seen in Fig. 6, it preserves probability mass at the modes of the gray densities.

B. Prediction of Pseudo Gaussian Densities

For static systems, we have shown at the beginning that a proper transformation $\underline{T}$ is obtained from (13). Simultaneously to the sensor models, many deterministic systems can also be linearized through a state space transformation. As mentioned before, a simultaneous linearization of stochastic system dynamics and stochastic sensor models is not possible in general. Through $\underline{T}$, linear system dynamics may even become nonlinear, while the sensor models turn to linear mappings. However, in many applications, a local prediction step for the state estimate is required at each node in a sensor network.

In order to obtain an approximate prediction of the transformed state $\underline{\boldsymbol{x}}_k^*$, either the density (15) or the system (14) can be considered. In the former case, [24] have shown how to predict exponentials of polynomials, which imply that (11) and (14) are polynomial mappings (or approximations). In the latter case, feedback and Carleman linearizations [29]–[31] can be applied. Carleman linearization techniques turn systems into bilinear systems, where the product of two Gaussian

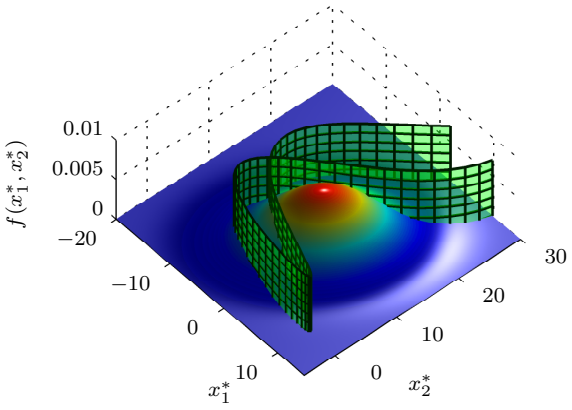


Figure 7. Result of CI algorithm in the transformed state space $\mathcal{S}^*$.

random vectors can easily be reapproximated by the first two moments.

In the following section, we employ a naïve sample-based approximation of the system dynamics: Let $\{\underline{\xi}_j^*\}_{j=1,\dots,W}$ be W samples of the current estimate $(\underline{\hat{x}}_k^*, \mathbf{C}_k^*)$, which can be computed randomly or deterministically. According to (10), the inverse $T_i^{-1}(\xi_{j,i}^*)$ is calculated for every component $\xi_{j,i}^*$. If T_i is only locally invertible on the subsets $\mathcal{S}_1^*, \mathcal{S}_2^*, \dots$, then T_i^{-1} is applied on each of these subsets. E.g., for $T_i(x) = x^2 = \underline{x}_i^*$, we obtain the samples $T_i^{-1}(\underline{\xi}_i^*) = \pm\sqrt{\underline{\xi}_i^*}$.[3] The obtained samples are then propagated through the system model (14), mapped back to $\mathcal{S}^*$, and approximated by their first two moments, in order to compute $(\underline{\hat{x}}_{k+1}^*, \mathbf{C}_{k+1}^*)$.

Of course, an approximation of the system model entails the risk that nonlinear dependencies between estimates are neglected that lead to a situation as depicted in Fig. 1. However, in order to counteract this issue and at the cost of higher computational demands, one can choose transformations to higher-order systems, where nonlinearities are almost negligible.

VI. Simulations

In this section, we will exemplify and discuss the presented approach by means of two simulations. The first simulation considers a one-dimensional system with a complicated measurement model. In the second example, an object is tracked by means of three nodes equipped with distance sensors.

A. One-dimensional System

At first, we confine ourselves to a static state x with value 3. It is observed by five nodes with sensor model

$$\hat{\underline{z}} = \begin{bmatrix} 10\sin(\boldsymbol{x}+1) \\ \boldsymbol{x}^2 \end{bmatrix} + \underline{\boldsymbol{v}} \ ,$$

[3]This doubles the number of samples.

where $\underline{\boldsymbol{v}}$ is a zero-mean noise term with covariance matrix $\mathbf{C}_v = 3 \cdot \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}$. The state is transformed according to

$$\underline{T}(x) = [10 \cdot \sin(x+1),\ x^2]^{\mathrm{T}} \ .$$

Fig. 7 shows the estimated pseudo Gaussian density after four time steps, where the nodes have communicated their results at every time step by utilizing the CI algorithm. The densities in Fig. 8 correspond to the manifold. In particular, the pseudo Gaussian density in Fig. 7 yields the green EMD in Fig. 8. The blue density characterizes the optimal result, where all

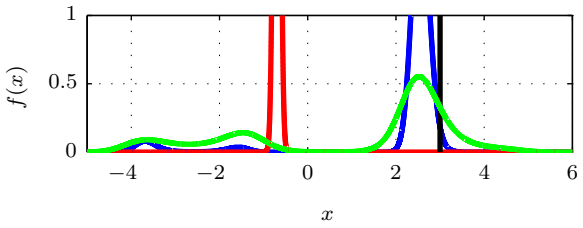


Figure 8. Densities in original space. The black line marks the true state. Green line: CI fusion result. Blue line: Optimal fusion result. Red line: EKF.

dependencies have been taken into account. The red density is obtained by linearizing the measurement mapping directly, i.e., applying the extended Kalman filter (EKF). Although the correlations have been incorporated, the result is strongly biased. Fig. 9 shows the root-mean-squared-error (RMSE) over 200 Monte Carlo runs, where 10 fusion steps have been performed in each run. The blue line corresponds to

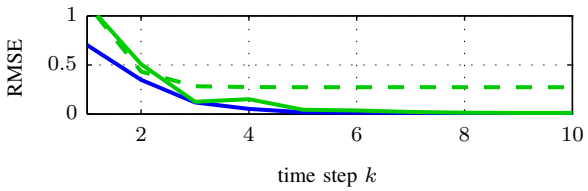


Figure 9. Static system: RMSE.

the optimal Bayesian fusion. For the solid green line, the CI algorithm has been applied to the pseudo Gaussian densities. The dashed green line also bases on pseudo Gaussian densities with the difference that dependencies have been ignored.

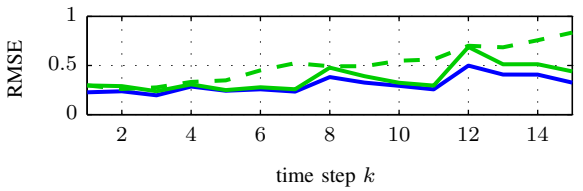
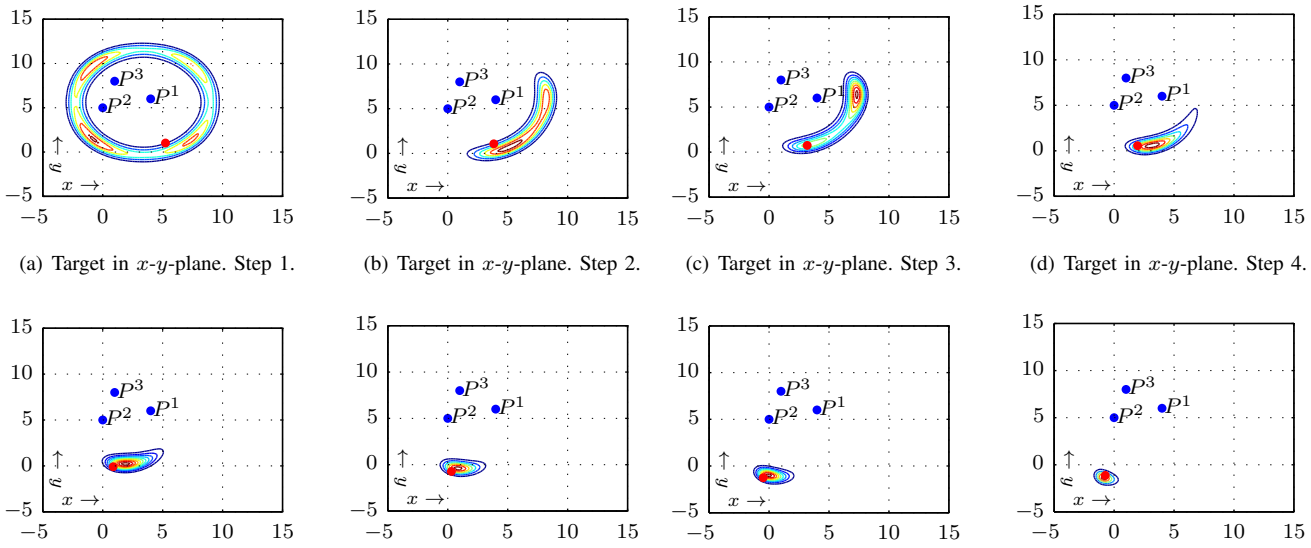


Figure 10. Dynamic system: RMSE.

In a second simulation, a dynamic evolution of x by means of the model

$$\boldsymbol{x}_{k+1} = 1.1 \cdot \boldsymbol{x}_k + \boldsymbol{w}_k$$



(a) Target in x-y-plane. Step 1. (b) Target in x-y-plane. Step 2. (c) Target in x-y-plane. Step 3. (d) Target in x-y-plane. Step 4.

(e) Target in x-y-plane. Step 5. (f) Target in x-y-plane. Step 6. (g) Target in x-y-plane. Step 7. (h) Target in x-y-plane. Step 8.

Figure 11. Probability density contours at different time instants calculated at sensor node P^1. The CI algorithm is applied in $\mathcal{S}^*$ in order to fuse the local estimates at each step. The red dot marks the true position of the target.

is considered, where $\boldsymbol{w}_k$ has zero-mean and variance 2. The prediction is sample-based as explained in Section V-B. Fig. 10 again illustrates the RMSE.

B. Localization in a Sensor Network

As a multi-dimensional example, we consider the tracking of a moving target by means of three distance sensors, which are located at $P^1 = [1, 8]^{\mathrm{T}}$, $P^2 = [0, 5]^{\mathrm{T}}$, and $P^3 = [4, 6]^{\mathrm{T}}$. The sensor models are

$$\begin{aligned}\hat{\underline{z}}_k^i &= (\boldsymbol{x}_k - P_x^i)^2 + (\boldsymbol{y}_k - P_y^i)^2 + \boldsymbol{v}_k \\ &= \mathbf{H}^* \, \underline{\boldsymbol{x}}_k^* + (P_x^i)^2 + (P_y^i)^2 + \boldsymbol{v}_k\end{aligned}$$

with the transformed measurement mapping

$$\mathbf{H}^* = [-2P_x^i, 1, -2P_y^i, 1]$$

and the transformed state $\underline{\boldsymbol{x}}_k^* = [\boldsymbol{x}_k, \boldsymbol{x}_k^2, \boldsymbol{y}_k, \boldsymbol{y}_k^2]^{\mathrm{T}}$. The zero-mean measurement noise $\boldsymbol{v}_k$ has variance 16. The target moves linearly with constant velocity, i.e.,

$$\begin{bmatrix}\boldsymbol{x}_{k+1} \\ \boldsymbol{y}_{k+1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{x}_k \\ \boldsymbol{y}_k\end{bmatrix} + \left(\begin{bmatrix}-1 \\ -0.5\end{bmatrix} + \underline{\boldsymbol{w}}_k\right) \ ,$$

where the input noise $\underline{\boldsymbol{w}}_k$ has the covariance matrix $\mathrm{diag}([0.2, 0.2])$. In this simulation, we have calculated the first two moments of $\underline{x}_{k+1}^*$ analytically. For example, since the model is linear, the second component of $\underline{x}_{k+1}^*$ is of the form

$$\boldsymbol{x}_{k+1,2}^* = \boldsymbol{x}_{k+1}^2 = \left(a \cdot \boldsymbol{x}_k + b \cdot (\hat{u}_k + \boldsymbol{w}_k)\right)^2 \ ,$$

where $\hat{u}_k$ and $\boldsymbol{w}_k$ denote a control input and input noise, respectively. For this quadratic mapping, the moments can directly be computed. In Fig. 11, the densities in the original state space are depicted over eight time steps. More precisely, they belong to the local estimate in sensor node 1, where the nodes have communicated their estimates at every time step by means of the CI fusion algorithm.

VII. Conclusions and Outlook

Dealing with unknown dependencies in nonlinear estimation problems is particularly challenging, since possible dependency structures cannot easily be parameterized or classified. This is in contrast to linear estimation fusion, where solely correlations between estimates need to be considered. In this paper, we have addressed this problem by employing a state space transformation that turns every operation on the state estimates to a linear mapping, i.e., fusion, measurement update, and prediction. For the linearization of the former two operations, such a state space transformation can in general be found. Especially for stochastic system dynamics, the prediction step has to be approximated in the transformed state space. In this case, state space transformations are desirable that alleviate linearization errors, so that only weak nonlinear dependencies between estimates can arise. In this regard, it appears to be very promising to further investigate the direction of [30] and other linearization techniques based on state transformations.

We generally believe that for suboptimal nonlinear estimation, additional knowledge on the possible dependency structures is inevitable. Otherwise, very arbitrary fusion results have to be considered, for which a "conservative" bound has to be determined, and it is actually an open question what conservativeness means for non-Gaussian probability densities. Fortunately, the possible dependencies between local estimate are determined through the underlying system, sensor, and fusion models.

By means of the state space transformation, we have achieved that, again, dependencies can uniquely be parameterized by correlations. If they are known, the Bar-Shalom/Campo rule can be deployed for fusion. If they are unknown, the CI algorithm provides suboptimal fused estimates. Moreover, we can employ the notion of covariance consistency in the transformed space. In this respect, we have shown that CI directly correspond to the EMD fusion model in the original state state. Possibly, this can be regarded as an indication that EMDs are related to linear dependency structures in a different state space. Also, we think that these transformations reveal a direction to define consistency in nonlinear estimation, which should be a topic in prospective research. However, consistency finally depends on the perspective of the user, i.e., the parameters of interest.

VIII. Acknowledgment

This work was partially supported by the German Research Foundation (DFG) within the Research Training Group GRK 1194 "Self-organizing Sensor-Actuator-Networks".

References

[1] S. Grime and H. Durrant-Whyte, "Data Fusion in Decentralized Sensor Networks," *Control Engineering Practice*, vol. 2, no. 5, pp. 849–863, 1994.

[2] M. E. Liggins, C.-Y. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulis, "Distributed Fusion Architectures and Algorithms for Target Tracking," in *Proceedings of the IEEE*, vol. 85, no. 1, 1997.

[3] M. E. Liggins, D. L. Hall, and J. Llinas, Eds., *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed., ser. The Electrical Engineering and Applied Signal Processing Series. CRC Press, 2009.

[4] D. Fränken and A. Hüpper, "Improved Fast Covariance Intersection for Distributed Data Fusion," in *Proceedings of the 8th International Conference on Information Fusion (Fusion 2005)*, Philadelphia, Pennsylvania, USA, Jul. 2005.

[5] C.-Y. Chong and S. Mori, "Graphical Models for Nonlinear Distributed Estimation," in *Proceedings of the 7th International Conference on Information Fusion (Fusion 2004)*, Stockholm, Sweden, Jun. 2004.

[6] Y. Wang and X. R. Li, "Distributed Estimation Fusion under Unknown Cross-correlation: An Analytic Center Approach ," in *Proceedings of the 13th International Conference on Information Fusion (Fusion 2010)*, Edinburgh, United Kingdom, Jul. 2010.

[7] J. K. Uhlmann, "General Data Fusion for Estimates with Unknown Cross Covariances," in *Proceedings of SPIE Aerosense Conference*, vol. 2755, 1996, pp. 536–547.

[8] S. J. Julier and J. K. Uhlmann, "A Non-divergent Estimation Algorithm in the Presence of Unknown Correlations," in *Proceedings of the IEEE American Control Conference (ACC 1997)*, vol. 4, Albuquerque, New Mexico, USA, Jun. 1997, pp. 2369–2373.

[9] S. J. Julier and J. K. Uhlmann, *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed. CRC Press, 2009, ch. General Decentralized Data Fusion with Covariance Intersection, pp. 319–343.

[10] U. D. Hanebeck and K. Briechle, "New Results for Stochastic Prediction and Filtering with Unknown Correlations," in *Proceedings of the 2001 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2001)*, Baden–Baden, Germany, Aug. 2001, pp. 147–152.

[11] U. D. Hanebeck, K. Briechle, and J. Horn, "A Tight Bound for the Joint Covariance of Two Random Vectors with Unknown but Constrained Cross–Correlation," in *Proceedings of the 2001 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2001)*, Baden–Baden, Germany, Aug. 2001, pp. 85–90.

[12] M. B. Hurley, "An Information Theoretic Justification for Covariance Intersection and its Generalization," in *Proceedings of the 5th International Conference on Information Fusion (Fusion 2002)*, Annapolis, Maryland, USA, Jul. 2002.

[13] Y. Wang and X. R. Li, "A Fast and Fault-Tolerant Convex Combination Fusion Algorithm under Unknown Cross-Correlation," in *Proceedings of 12th International Conference on Information Fusion (Fusion 2009)*, Seattle, Washington, USA, Jul. 2009.

[14] C.-Y. Chong and S. Mori, "Convex Combination and Covariance Intersection Algorithms in Distributed Fusion," in *Proceedings of the 4th International Conference on Information Fusion (Fusion 2001)*, Montréal, Québec, Canada, Aug. 2001.

[15] J. Sijs, M. Lazar, and P. P. J. van den Bosch, "State-Fusion with Unknown Correlation: Ellipsoidal Intersection," in *Proceedings of the 2010 American Control Conference (ACC 2010)*, Baltimore, Maryland, USA, Jun. 2010.

[16] S. J. Julier, T. Bailey, and J. K. Uhlmann, "Using Exponential Mixture Models for Suboptimal Distributed Data Fusion," in *IEEE Nonlinear Statistical Signal Processing Workshop (NSSPW 2006)*, Cambridge, United Kingdom, Sep. 2006, pp. 160–163.

[17] S. J. Julier, "An Empirical Study into the Use of Chernoff Information for Robust, Distributed Fusion of Gaussian Mixture Models," in *Proceedings of the 9th International Conference on Information Fusion (Fusion 2006)*, Florence, Italy, Jul. 2006.

[18] S. J. Julier, "Fusion without Independence (keynote abstract)," in *Proceedings of the IET Seminar on Tracking and Data Fusion: Algorithms and Applications*, Birmingham, United Kingdom, Apr. 2008, pp. 1–5.

[19] F. Sawo, D. Brunn, and U. D. Hanebeck, "Parameterized Joint Densities with Gaussian and Gaussian Mixture Marginals," in *Proceedings of the 9th International Conference on Information Fusion (Fusion 2006)*, Florence, Italy, Jul. 2006.

[20] Y. Bar-Shalom and L. Campo, "The Effect of the Common Process Noise on the Two-Sensor Fused-Track Covariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 6, pp. 803–805, Nov. 1986.

[21] S. J. Julier, "Estimating and Exploiting the Degree of Independent Information in Distributed Data Fusion," in *Proceedings of the 12th International Conference on Information Fusion (Fusion 2009)*, Seattle, Washington, USA, Jul. 2009.

[22] S. Reece and S. Roberts, "Robust, Low-Bandwidth, Multi-Vehicle Mapping," in *Proceedings of the 8th International Conference on Information Fusion (Fusion 2005)*, vol. 2, Philadelphia, Pennsylvania, USA, Jul. 2005.

[23] R. P. S. Mahler, "Optimal/Robust Distributed Data Fusion: A Unified Approach," in *Proceedings of SPIE*, vol. 4052, 2000.

[24] B. Tonkes and A. D. Blair, "Decentralised Data Fusion with Exponentials of Polynomials," in *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2007)*, San Diego, California, USA, Oct. 2007.

[25] D. Clark, S. J. Julier, R. P. S. Mahler, and B. Ristic, "Robust Multi-Object Sensor Fusion with Unknown Correlations," in *Sensor Signal Processing for Defence (SSPD 2010)*, London, United Kingdom, Sep. 2010.

[26] U. D. Hanebeck, "Optimal Filtering of Nonlinear Systems Based on Pseudo Gaussian Densities," in *Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003)*, Rotterdam, Netherlands, Aug. 2003, pp. 331–336.

[27] U. D. Hanebeck, "Recursive Nonlinear Set–Theoretic Estimation Based on Pseudo–Ellipsoids," in *Proceedings of the 2001 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2001)*, Baden–Baden, Germany, Aug. 2001, pp. 159–164.

[28] T. Lefebvre, H. Bruyninckx, and J. de Schutter, *Nonlinear Kalman Filtering for Force-Controlled Robot Tasks*, ser. Springer Tracts in Advanced Robotics. Springer, 2005, vol. 19.

[29] K. Kowalski and W.-H. Steeb, *Nonlinear Dynamical Systems and Carleman Linearization*. World Scientific Publishing Company, 1991.

[30] A. Germani, C. Manes, and P. Palumbo, "Polynomial Extended Kalman Filter," *IEEE Transactions on Automatic Control*, vol. 50, no. 12, pp. 2059–2064, Dec. 2005.

[31] A. Germani and C. Manes, "The Polynomial Extended Kalman Filter as an Exponential Observer for Nonlinear Discrete-Time Systems," in *Proceedings of the 47nd IEEE Conference on Decision and Control (CDC 2008)*, Cancún, Mexico, Dec. 2008.