State fusion with unknown correlation : ellipsoidal intersection

State fusion with unknown correlation: Ellipsoidal intersection

J. Sijs, *Student Member, IEEE,* M. Lazar, *Member, IEEE,* P.P.J.v.d. Bosch, *Member, IEEE.*

Abstract— Some crucial challenges of estimation over sensor networks are reaching consensus on the estimates of different systems in the network and separating the mutual information of two estimates from their exclusive information. Current fusion methods of two estimates tend to bypass the mutual information and directly optimize the fused estimate. Moreover, both the mean and covariance of the fused estimate are fully determined by optimizing the covariance only. In contrast to that, this paper proposes a novel fusion method in which the mutual information results in an additional estimate, which defines a mutual mean and covariance. Both variables are derived from the two initial estimates. The mutual covariance is used to optimize the fused covariance, while the mutual mean optimizes the fused mean. An example of decentralized state estimation, where the proposed fusion method is applied, shows a reduction in estimation error compared to the existing alternatives.

Index Terms— State fusion, decentralized state estimation.

I. INTRODUCTION

A current trend in networked systems is to disperse estimation algorithms among the different subsystems, i.e., nodes, rather than running a centralized algorithm. The main advantage of such an approach is that communication and/or computation requirements of a single node decreases while robustness to node-failure increases. To profit from these advantages in the case of decentralized state-estimators, two challenges are first to be solved. One of these challenges is on the fusion of estimates, which is the focus of this research.

We consider estimation algorithms with a probability density function that is described by a Gaussian. A well known example of such an algorithm is the Kalman filter [1]. When designing a decentralized version of this estimator that is suitable for a sensor network, a widely accepted solution is to perform a local the Kalman filter that uses the node's measurement [2]–[6]. To improve the node's local estimate, data can be exchanged with neighboring nodes. An example where nodes exchange their local measurement can be found in [3]. A drawback when exchanging measurements is that each node has access to a different set of measurements. Therefore, the estimation-result will differ per node, which is the first challenge of decentralized state-estimation algorithms. This can be solved when the mean of the estimated state-vector is exchanged as well. A consensus step between the local and the received estimated state-vectors ensures that the mean of each node converges to the

same value, as presented in [4]. However, in the consensus-step accurate estimates are treated with an equal importance as inaccurate ones. As such, the estimation-error of accurate estimates will *increase* after the consensus-step. Hence, the second challenge in decentralized estimation algorithms is to combine two estimates into one "fused estimate". The main difficulty is to cope with the correlation between two estimates that occurs when they are (partly) based on mutual information. In case mutual information is treated as exclusive information, the estimates become "over-confident" and incorrect.

To that extent, in this paper we propose a novel fusion method of two estimates with unknown mutual information. This assumption is required for networked systems as it is difficult to keep track of mutual information, due to the large amount of interaction between the nodes. Firstly, a method is derived to calculate the fused estimate. Therein, the effect of mutual information is explicitly taken into account in terms of a mutual mean and covariance. Secondly, a novel method that estimates the mutual mean and covariance, by assuming a maximum effect of the mutual information, is proposed. A benchmark case study, i.e., temperature estimation, is employed to illustrate the improvement of current decentralized state-estimators when using the proposed fusion method in comparison with other existing fusion methods.

II. PRELIMINARIES

$\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{Z}$ and $\mathbb{Z}_+$ define the set of real numbers, non-negative real numbers, integer numbers and non-negative integer numbers, respectively. For any $\mathscr{C} \subset \mathbb{R}$, let $\mathbb{Z}_{\mathscr{C}} := \{c \in \mathbb{Z} | c \in \mathscr{C}\}$. The notation 0 is used to denote either the null-vector or null-matrix. Its size will become clear from the context. The transpose, inverse, determinant and trace of a matrix $A \in \mathbb{R}^{n \times n}$ are denoted as $A^\top$, A^{-1}, $|A|$ and $\text{tr}(A)$, respectively. Further, $[A]_{ij} \in \mathbb{R}$ denotes the element on the ith row and j^{th} column of A. Given that $A, B \in \mathbb{R}^{n \times n}$ are positive definite, denoted with $A \succ 0$ and $B \succ 0$, then $A \succ B$ denotes $A - B \succ 0$. $A \succeq 0$ denotes that A is positive semi-definite. Given the square matrix $A \in \mathbb{R}^{n \times n}$, let $v_q(A) \in \mathbb{R}^n$ and $\lambda_q(A) \in \mathbb{R}$ denote the qth eigenvector and eigenvalue, respectively. If $v_q(A)$ and $\lambda_q(A)$ contain only real values, for all $q \in \mathbb{Z}_{[1,n]}$, then $A = S_A D_A S_A^{-1}$ denotes the Jordan decomposition of A, where:

$$
\begin{aligned}
S_A &:= (v_1(A), v_2(A), \dots, v_n(A)), \\
D_A &:= \text{diag}(\lambda_1(A), \lambda_2(A), \dots, \lambda_n(A)).
\end{aligned} \tag{1}
$$

The probability density function (PDF) of a random vector $x \in \mathbb{R}^n$ is denoted as $p(x)$. The Gaussian function (shortly noted as Gaussian) is denoted with $G(x, \mu, P)$, for some $x, u \in$

J. Sijs is with TNO Science and Industry, P.O. Box 155, 2600 AD Delft, The Netherlands, E-mail: `joris.sijs@tno.nl`.

M. Lazar and P. v. d.Bosch are with Eindhoven University of Technology, Eindhoven, The Netherlands, E-mail: `m.lazar@tue.nl`, `p.p.j.v.d.bosch@tue.nl`.

$\mathbb{R}^n$ and $P \in \mathbb{R}^{n \times n}$. If $p(x) = G(x, \mu, P)$, then by definition it holds that $E[x] = \mu$ and $cov(x) = P$. Moreover, $P \succeq 0$ is a symmetric matrix. Any Gaussian $G(x, \mu, P)$ can be represented by its sub-level-set $\varepsilon_{\mu,P}(x) \subset \mathbb{R}^n$, which is an ellipsoidal set defined as follows:

$$\varepsilon_{\mu,P}(x) := \left\{ x \middle| (x - \mu)^T P^{-1} (x - \mu) \leq 1 \right\}. \tag{2}$$

An example of $\varepsilon_{\mu,P}(x)$ is graphically depicted in Figure 1.
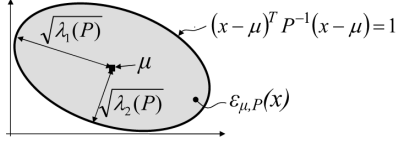

Fig. 1. Representation of the Gaussian $G(x, \mu, P)$ by its sub-level-set $\varepsilon_{\mu,P}(x)$. The covariance-matrix P is chosen such that $\lambda_1(P) < \lambda_2(P)$.

III. PROBLEM FORMULATION

Let us assume a process which is observed by a network of sensor-nodes. The goal is to estimate the state-vector $x \in \mathbb{R}^n$ of the process. To that extent, each node i calculates a PDF of x at predefined sample instants k, denoted with $p_i(x[k])$. We assume that all these PDFs are Gaussians with mean $x_i[k] \in \mathbb{R}^n$ and covariance $P_i[k] \in \mathbb{R}^{n \times n}$, i.e.,

$$p_i(x[k]) := G(x, x_i[k], P_i[k]). \tag{3}$$

To improve a node's local estimate, reach consensus on the estimates at the different nodes and be able to deal with correlations, each node shares its local estimate with its neighboring nodes. Therefore, if the neighboring nodes of node i are collected in the set N_i, then node i receives the multiple estimates $p_j(x[k])$, for all $j \in N_i$. Fusion of the received estimates with $p_i(x[k])$ is done in a fusion algorithm, which results in the fused estimate at node i, denoted with $p_{i_f}(x[k])$. The algorithm performs a fusion of *two* estimates sequentially, N_i times. Each sequence incorporates the next received estimate. This mode of operation is graphically depicted in Figure 2.
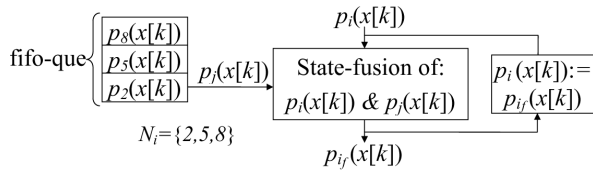

Fig. 2. The fusion function is a part of the algorithm to fuse the local estimate $p_i(x[k])$ with the receiving ones $p_j(x[k])$, for all $j \in N_i$.

The goal of this paper is to design a fusion method that combines a local estimate $p_i(x[k])$ with an arbitrary received estimate $p_j(x[k])$, $j \in N_i$, into one improved estimate $p_{i_f}(x[k])$. It is assumed that the mutual information, i.e., correlation, between the two estimates is unknown. Furthermore, it is assumed that all estimates are described by a Gaussian, i.e., $p_j(x[k]) = G(x, x_j[k], P_j[k])$ and $p_{i_f}(x[k]) = G(x, x_{i_f}[k], P_{i_f}[k])$. An example of $p_i(x[k])$ and $p_j(x[k])$, with their corresponding sub-level-sets, is shown in Figure 3(a).

A. Related work on state-fusion

Current fusion methods of two estimates with unknown correlation are mostly based on Covariance Intersection (CI) [6]–[8]. This method defines that, for a certain $\omega \in [0,1]$, the fused estimates are a convex combination of the parameters in $p_i(x[k])$ and $p_j(x[k])$, i.e.,

$$x_{i_f}[k] = \omega x_i[k] + (1 - \omega) x_j[k], \; P_{i_f}[k] = \omega P_i[k] + (1 - \omega) P_j[k].$$

The optimal value of ω is found by minimizing $\text{tr}(P_{i_f})$, due to which ω only depends on the trace (or determinant) of P_i and P_j. Therefore, both x_{i_f} and P_{i_f} depend only on the limited information of $\text{tr}(P_i)$ and $\text{tr}(P_j)$. Moreover, for any amount of mutual information it is proven that $P_{i_f} \preceq P_i$ and $P_{i_f} \preceq P_j$ [9], which yields the property $\varepsilon_{0,P_{i_f}}(x) \subseteq \varepsilon_{0,P_i}(x) \cap \varepsilon_{0,P_j}(x)$. Therein it was also proven that this property does not hold for CI, as it is illustrated in Figure 3(b).
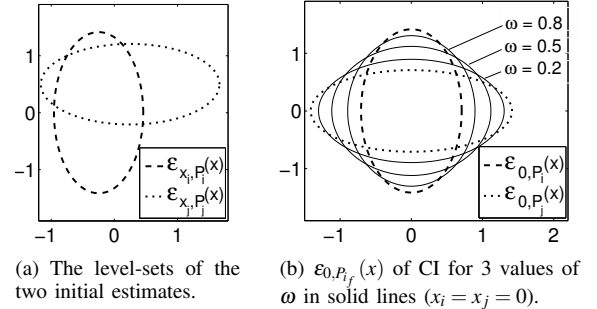

(a) The level-sets of the two initial estimates.

(b) $\varepsilon_{0,P_{i_f}}(x)$ of CI for 3 values of ω in solid lines ($x_i = x_j = 0$).

Fig. 3. An example of two estimates $p_i(x[k])$ and $p_j(x[k])$.

To ensure this property the scalars ω and $(1 - \omega)$ are replaced with some matrices W_i and W_j, respectively, as proposed in [9]. However, both matrices are computed with an iterative algorithm to minimize $\text{tr}(P_{i_f})$, which requires significant processing power. Also, x_{i_f} is fully defined when optimizing $\text{tr}(P_{i_f})$, instead of formulating an optimization depending on x_{i_f}. A different (heuristic) state-fusion method, in which P_{i_f} is optimized, was proposed in [10].

To conclude, the objective of all current fusion methods is to optimize P_{i_f} or its trace/determinant. However, optimizing the fusion result is not a desirable objective. Instead, the actual problem is determining the effect of the mutual information to such an extent that $p_i(x[k])$ is updated only with *exclusive* information from $p_j(x[k])$. Additionally, current methods do not optimize x_{i_f}. This indicates that the fused mean is of less importance than the fused covariance is. This is not suitable for most control methodologies as they rely on x rather than on P. To solve these issues, the next section presents a novel fusion method centered around an estimation of the state-vector based on the mutual information.

IV. ELLIPSOIDAL INTERSECTION

The goal of Ellipsoidal Intersection (EI) is to use only exclusive information of $p_j(x)$ to update $p_i(x)$ by separating their mutual information first. To that extent a new estimate of x is defined, which is based on the mutual information of $p_i(x)$ and $p_j(x)$ only. The PDF of this new estimate

is modeled as a Gaussian with a mutual mean γ and a mutual covariance Γ. We will first show how x_{i_f} and P_{i_f} are determined in case the mutual information, i.e., γ and Γ, is known. After that a method is presented to estimate both Γ and γ. As correlation is unknown, we will estimate Γ and γ by assuming a maximum effect of the mutual information.

A. Fusion with mutual information

The estimates $p_i(x)$ and $p_j(x)$ are based on information which could be (partly) mutual. Let us denote their mutual information as $d_\gamma \in \mathbb{R}^n$ and their exclusive information as $d_i \in \mathbb{R}^n$ and $d_j \in \mathbb{R}^n$ for $p_i(x)$ and $p_j(x)$, respectively. With this, $p_i(x)$ and $p_j(x)$ can be rewritten as a conditional PDF, which can be further derived by applying Bayes' rule [11]:

$$p_i(x) := p(x|d_\gamma, d_i) = \frac{p(x|d_\gamma)p(d_i|x)}{\int_{-\infty}^{\infty} p(x|d_\gamma)p(d_i|x)dx}, \quad (4a)$$

$$p_j(x) := p(x|d_\gamma, d_j) = \frac{p(x|d_\gamma)p(d_j|x)}{\int_{-\infty}^{\infty} p(x|d_\gamma)p(d_j|x)dx}. \quad (4b)$$

The exclusive information within $p_j(x)$ is d_j. Therefore, the fused PDF becomes $p_{i_f}(x) := p(x|d_\gamma, d_i, d_j)$, i.e.,

$$p_{i_f}(x) = \frac{p(x|d_\gamma, d_i)p(d_j|x)}{\int_{-\infty}^{\infty} p(x|d_\gamma, d_i)p(d_j|x)dx}, \quad (5a)$$

$$= \frac{p_i(x)p(d_j|x)}{\int_{-\infty}^{\infty} p_i(x)p(d_j|x)dx}. \quad (5b)$$

As the PDFs of the initial estimates are Gaussian, let us model $p(d_j|x)$ and $p(x|d_\gamma)$ as Gaussians as well, i.e., $p(d_j|x) := G(\mu_j, x, U_j)$ and $p(x|d_\gamma) := G(x, \gamma, \Gamma)$, for some $\mu_j, \gamma \in \mathbb{R}^n$ and $U_j, \Gamma \in \mathbb{R}^{n \times n}$.

Proposition IV.1 [1], [12] *Let there exist two Gaussian PDFs of random vectors $x, w \in \mathbb{R}^n$ defined with $v \in \mathbb{R}^n$ and $W, V \in \mathbb{R}^{n \times n}$; $p(x|v) = G(x, v, V)$ and $p(w|x) = G(w, x, W)$. If $p(x|v)$ and $p(w|x)$ are uncorrelated, then they satisfy:*

$$p(x|v, w) = \frac{p(x|v)p(w|x)}{\int_{-\infty}^{\infty} p(x|v)p(w|x)dx} = G(x, z, Z),$$

where $Z = \left(V^{-1} + W^{-1}\right)^{-1}, z = Z\left(v^{-1}v + W^{-1}w\right).$

Due to the fact that d_j is exclusive information, it follows that $p(d_j|x)$ and $p_i(x)$ are uncorrelated. Therefore, applying Proposition IV.1 in (5b), by substituting $G(x, v, V) = G(x, x_i, P_i)$, $G(w, x, W) = G(\mu_j, x, U_j)$ and $G(x, z, Z) = G(x, x_{i_f}, P_{i_f})$, gives that:

$$P_{i_f} = \left(P_i^{-1} + U_j^{-1}\right)^{-1}, \quad x_{i_f} = P_{i_f}\left(P_i^{-1}x_i + U_j^{-1}\mu_j\right). \quad (6)$$

In case we assume that γ and Γ are known, then μ_j and U_j are derived by applying Proposition IV.1 in (4b). In that case the substitution implies $G(x, v, V) = G(x, \gamma, \Gamma)$, $G(w, x, W) = G(\mu_j, x, U_j)$ and $G(x, z, Z) = G(x, x_j, P_j)$ and results in:

$$P_j = \left(\Gamma^{-1} + U_j^{-1}\right)^{-1}, \quad x_j = P_j\left(\Gamma^{-1}\gamma + U_j^{-1}\mu_j\right), \quad (7a)$$

$$\Rightarrow U_j^{-1} = P_j^{-1} - \Gamma^{-1}, \quad \mu_j = U_j\left(P_j^{-1}x_j - \Gamma^{-1}\gamma\right). \quad (7b)$$

Substituting the results of (7b) into equation (6) gives the fused mean x_{i_f} and covariance P_{i_f}, which now depend on the mutual mean γ and mutual covariance Γ, i.e.,

$$P_{i_f} = \left(P_i^{-1} + P_j^{-1} - \Gamma^{-1}\right)^{-1},$$

$$x_{i_f} = P_{i_f}\left(P_i^{-1}x_i + P_j^{-1}x_j - \Gamma^{-1}\gamma\right). \quad (8)$$

Notice the difference of these fusion equations compared to the ones of CI. In the case of CI the expression resembles to an *agreement*, i.e., it is a convex combination of $p_i(x)$ and $p_j(x)$. In contrast to that, equation (8) is an *update* of $p_i(x)$ with the exclusive information of $p_j(x)$.

Equation (8) shows how x_{i_f} and P_{i_f} are determined if the mutual mean and covariance, i.e., γ and Γ, are known. The next step is to estimate their corresponding values.

B. Mutual covariance

The goal is now to find a value for Γ such that the mutual information between $p_i(x)$ and $p_j(x)$ is maximized. This means that the modeled accuracy of the estimation due to the mutual information only, i.e., $\lambda_q(\Gamma)$ for all $q \in \mathbb{Z}_{[1,n]}$, is as "small" as possible. However, notice that U_j is a covariance matrix, i.e., $U_j \succ 0$. Therefore, it should satisfy $U_j^{-1} \succeq 0$, which, if applied in (7b), gives that $\Gamma \succeq P_j$. Similarly $\Gamma \succeq P_i$ must also hold. Let $\varepsilon_{0,P_i}(x)$, $\varepsilon_{0,P_j}(x)$ and $\varepsilon_{0,\Gamma}(x)$ denote the ellipsoidal sub-level-sets that correspond to these three covariances. Then, the above conditions are attained if $\varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x) \subseteq \varepsilon_{0,\Gamma}(x)$.

Definition IV.2 *Let $p_i(x)$ and $p_j(x)$ be given. If their mutual information is assumed to be maximum, then their mutual covariance is defined as $\Gamma_{\max} := \arg\min_{\Upsilon \in \mathbb{R}^{n \times n}} \sum_{q=1}^{n} \lambda_q(\Upsilon)$, subject to the condition $\varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x) \subseteq \varepsilon_{0,\Upsilon}(x)$.*

Definition IV.3 [13] *Let $\mathscr{C} \subset \mathbb{R}^n$ be a bounded set and let $\Gamma_L \in \mathbb{R}^{n \times n}$ be defined as:*

$$\Gamma_L := \arg\min_{\Upsilon \in \mathbb{R}^{n \times n}} \quad \log|\Upsilon| \quad (9a)$$

$$\text{subject to} \quad x^\top \Upsilon^{-1} x \leq 1, \quad \forall x \in \mathscr{C}. \quad (9b)$$

Then $\varepsilon_{0,\Gamma_L}(x)$ is the Löwner−John ellipsoid (LJE) of $\mathscr{C}$.

Theorem IV.4 *Let $p_i(x)$ and $p_j(x)$ be given. If we substitute $\mathscr{C} = \varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x)$ in Definition IV.3, then $\Gamma_{\max} = \Gamma_L$.*

The proof is presented in Appendix A. Notice that Theorem IV.4, together with Definition IV.2 and Definition IV.3, yields $\Gamma = \Gamma_{\max} = \Gamma_L$. Before deriving an explicit solution for Γ, let us first prove that the LJE scales and rotates linearly in case of a transformation on its vector-space.

Lemma IV.5 *Let $A, B, \Gamma \succ 0$, the rotation matrix S (satisfying $S = S^{-\top}$) and the diagonal matrix $D \succ 0$ be given. Let x be transformed into $\hat{x} := DSx$ and let $\hat{A} := DSAS^{-1}D$, $\hat{B} := DSBS^{-1}D$ and $\hat{\Gamma} := DS\Gamma S^{-1}D$. If $\varepsilon_{0,\hat{\Gamma}}(\hat{x})$ is the LJE of $\varepsilon_{0,\hat{A}}(\hat{x}) \cup \varepsilon_{0,\hat{B}}(\hat{x})$, then $\varepsilon_{0,\Gamma}(x)$ is the LJE of $\varepsilon_{0,A}(x) \cup \varepsilon_{0,B}(x)$.*

The proof is presented in Appendix B. To minimize processing demand, we focus on deriving an explicit solution that corresponds to finding the LJE. As such, the next theorem presents the solution of the mutual covariance Γ from two initial PDFs $p_i(x)$ and $p_j(x)$. To that extent, let us define $S_i, S_j \in \mathbb{R}^{n \times n}$ and the diagonal matrices $D_i, D_j \in \mathbb{R}^{n \times n}$, such that:

$$P_i = S_i D_i S_i^{-1} \quad \text{and} \quad D_i^{-0.5} S_i^{-1} P_j S_i D_i^{-0.5} = S_j D_j S_j^{-1}. \quad (10)$$

Theorem IV.6 *Let $P_i, P_j \succ 0$ be given with their corresponding S_i, D_i, S_j and D_j according to (10). Let*

$$\Gamma := S_i D_i^{0.5} S_j D_\Gamma S_j^{-1} D_i^{0.5} S_i^{-1}, \quad (11a)$$

$$\text{with } [D_\Gamma]_{qr} := \begin{cases} \max([D_j]_{qr}, 1) & \text{if } q = r, \\ 0 & \text{if } q \neq r. \end{cases} \quad (11b)$$

Then $\varepsilon_{0,\Gamma}(x)$ is the LJE of $\varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x)$.

The proof is presented in Appendix C by first considering a transformation on x. The transformed matrices of P_i and P_j, denoted with $\hat{P}_i$ and $\hat{P}_j$, are diagonal and $\hat{P}_i = I$. Then the LJE of the transformed set $\varepsilon_{0,\hat{P}_i}(\hat{x}) \cup \varepsilon_{0,\hat{P}_j}(\hat{x})$, denoted with $\varepsilon_{0,D_\Gamma}(\hat{x})$, is determined and transformed from $\hat{x}$ back into x to define $\varepsilon_{0,\Gamma}(x)$. An example of this transformation, in case $P_i = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ and $P_j = \begin{pmatrix} 1/3 & 0 \\ 0 & 2 \end{pmatrix}$, is graphically depicted in Figure 4. This figure also shows the result of the mutual mean Γ and that $\varepsilon_{0,\hat{P}_i}(\hat{x})$ is the unit circle.
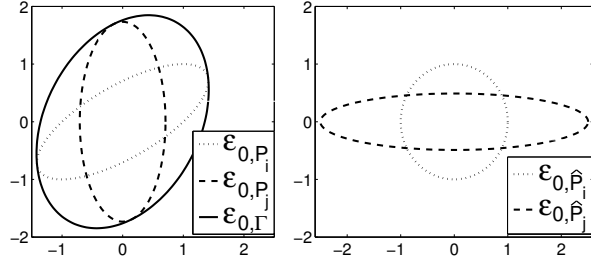

Fig. 4. The sub-level-sets $\varepsilon_{0,P_i}(x)$ and $\varepsilon_{0,P_j}(x)$ with their corresponding transformed ones, i.e., $\varepsilon_{0,\hat{P}_i}(\hat{x})$ and $\varepsilon_{0,\hat{P}_j}(\hat{x})$, and their LJE $\varepsilon_{0,\Gamma}(x)$.

Earlier it was stated that a proper fusion method must result in $P_{i_f} \preceq P_i$ and $P_{i_f} \preceq P_j$. Let us prove this property for EI before we continue with finding γ.

Lemma IV.7 *Let $\varepsilon_{0,\Gamma}(x)$ be defined as the LJE of $\varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x)$. Then $P_{i_f} \preceq P_i$ and $P_{i_f} \preceq P_j$.*

Proof: A property of the LJE is that $P_i \preceq \Gamma$ and $P_j \preceq \Gamma$. Notice that the latter inequality gives that $P_j^{-1} \succeq \Gamma^{-1}$ and thus $P_j^{-1} - \Gamma^{-1} \succeq 0$. Adding P_i^{-1} on both sides results in $P_i^{-1} + P_j^{-1} - \Gamma^{-1} \succeq P_i^{-1}$. From (8) it follows that the left hand side of this inequality equals $P_{i_f}^{-1}$. Therefore $P_{i_f}^{-1} \succeq P_i^{-1}$ and thus $P_{i_f} \preceq P_i$. $P_{i_f} \preceq P_j$ can be proven similarly. ∎

C. Mutual mean

The mutual mean γ represents an estimated mean of x on which both initial estimates $p_i(x)$ and $p_j(x)$ "agree on most". As such, we aim at deriving a cost-function $J(\alpha)$, whose minimum corresponds to the "most-agreed" value of x, i.e., $\gamma = \arg\min_{\alpha \in \mathbb{R}^n} J(\alpha)$. A standard method to find the most-agreed vector α of the two vectors x_i and x_j is to minimize the distance between x_i and α and between x_j and α. However, as x_i and x_j have a different accuracy each distance should be weighted accordingly. Hence, let us define $J(\alpha)$, for some $W_i, W_j \succ 0$, as follows:

$$J(\alpha) := (\alpha - x_i)^\top W_i (\alpha - x_i) + (\alpha - x_j)^\top W_j (\alpha - x_j). \quad (12)$$

Minimizing $J(\alpha)$ equals to finding α for which it holds that $\delta J(\alpha)/\delta\alpha = 0$. With Proposition 10.6.1 of [14], i.e., $\delta(a^\top A a)/\delta a = a^\top(A + A^\top)$ holds for any square matrix A and vector a of suitable dimensions, we can determine $\delta J(\alpha)/\delta\alpha$ and thus γ, i.e.,

$$\frac{\delta J}{\delta\gamma} = 2(\alpha - x_i)^\top W_i + 2(\alpha - x_j)^\top W_j, \quad (13a)$$

$$\Rightarrow \gamma = (W_i + W_j)^{-1}(W_i x_i + W_j x_j). \quad (13b)$$

The last step is to define the values for W_i and W_j. Let us start by choosing $W_i = U_j^{-1}$ and $W_j = U_i^{-1}$. These weights will result in a γ that will be closer to x_i than to x_j, if the ellipsoid ε_{0,U_i} is larger and the ellipsoid ε_{0,U_j} is smaller. Notice that this corresponds to a situation where x_j has more exclusive information compared to the exclusive information of x_i. Hence, node j has more unique information to update γ into x_j, as shown in (7a). An issue with $W_i = U_j^{-1}$ and $W_j = U_i^{-1}$ is that one cannot guarantee that $W_i, W_j \succ 0$ but only $W_i, W_j \succeq 0$. Therefore, we add ηI to both $U_i^{-1} = P_i^{-1} - \Gamma^{-1}$ and $U_i^{-1} = P_j^{-1} - \Gamma^{-1}$, for some $\eta > 0$. Hence, equation (13b) gives that:

$$\gamma := \left(P_i^{-1} + P_j^{-1} - 2\Gamma^{-1} + 2\eta I\right)^{-1} \times \\ \left(\left(P_j^{-1} - \Gamma^{-1} + \eta I\right)x_i + \left(P_i^{-1} - \Gamma^{-1} + \eta I\right)x_j\right). \quad (14)$$

To minimize the effect of η on γ, its value must be as small as possible. Therefore, with $H := P_i^{-1} + P_j^{-1} - 2\Gamma^{-1}$ and $\lambda_{0_+}(H) \in \mathbb{R}_+$ defined as the smallest non-zero eigenvalue of H, let us define η as follows:

$$\eta := \begin{cases} 0 & \text{if } |H| \neq 0 \\ c \ll \lambda_{0_+}(H) & \text{if } |H| = 0. \end{cases} \quad (15)$$

Now that both Γ and γ can be estimated, let us present a fusion example. Figure 5 shows the sub-level-sets of two initial estimates $p_i(x)$ and $p_j(x)$, of the estimate due to mutual information, i.e., $G(x, \gamma, \Gamma)$, and of their fused estimate $p_{i_f}(x)$. Here, $x_i = (1, -2)^\top$, $P_i = \begin{pmatrix} 3 & 0 \\ 0 & 0.4 \end{pmatrix}$, $x_j = (-2, -1)^\top$, $P_j = \begin{pmatrix} 2 & -0.8 \\ -0.8 & 1 \end{pmatrix}$. Figure 6 shows the result of fusion according to CI, with the same initial estimates. Notice that the latter one resembles more to mutual agreement.
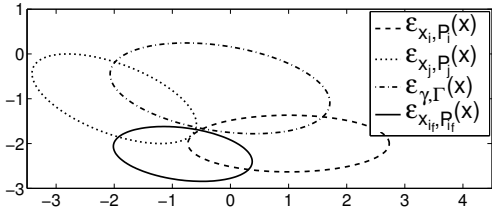
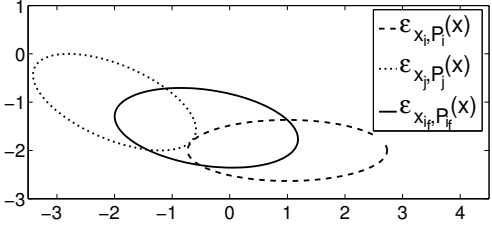Fig. 5. The two initial, mutual and fused estimates according to EI.

Fig. 6. The two initial and fused estimates according to CI.

V. SIMULATION CASE STUDY

In this section the fusion method EI is used in a decentralized Kalman filter and tested in terms of the achieved performance in estimation error. The benchmark process, graphically depicted in Figure 7, is the heat transfer of a bar which is starting at $300\ [K]$. The bar is divided into 11 segments and the temperature T_m of each segment m is estimated, i.e., the state-vector is $x = (T_1, T_2, \cdots, T_{11})^\top$. The process-model of heat transfer, in continuous time (t), is defined with $\delta T_1(t)/\delta t = 0$, $\delta T_{11}(t)/\delta t = 0$ and

$$500\dot{T}_6(t) = 2T_5(t) - 4T_6(t) + 2T_7(t) + 50,$$
$$500\dot{T}_m(t) = 2T_{m-1}(t) - 4T_m(t) + 2T_{m+1}(t), \ \forall m \in \mathbb{Z}_{[2,10]}\backslash\{6\}.$$

Notice that the bar is heated at the 6th segment with $50[W]$. The temperature distribution is estimated by a network of 5 nodes that are placed at the segments 2, 4, 6, 8 and 10, respectively. Each node is connected to its direct neighbor(s) only. A node i measures the temperature of its own segment, i.e., $C_2 = (0,0,0,1,0,0,0,0,0,0,0)$. A local estimation algorithm estimates the global state-vector at its local node i, denoted with x_i. Its local, discrete-time state-space model, derived from the continuous model, is described as:

$$x_i[k] = Ax_i[k-1] + w[k-1], \text{ with } p(w[k]) := G(w[k], 0, Q),$$
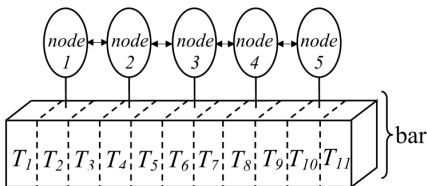$$y_i[k] = C_i x_i[k] + v_i[k], \qquad \text{with } p(v_i[k]) := G(v_i[k], 0, R_i).$$

Fig. 7. Heat transfer of a bar, measured with a networked system.

The sampling time is $10\ [s]$ and the model runs for 250 $[s]$ after which three different local estimation algorithms

are compared. The first algorithm, decentralized Ellipsoidal Intersection (DEI), contains two steps. The first one is a Kalman filter that uses the node's local measurement only. After this step, each node exchanges its estimate. The second step consists of the fusion algorithm EI, as presented in Section IV, in which c of (15) equals 10^{-10}. As such, the DEI algorithm at node i at sample instant k can be summarized as follows:

Step 1: local Kalman filter

$$M_i = AP_{i_f}[k-1]A^T + Q;$$
$$P_i = \left(M_i^{-1} + C_i^T R_i^{-1} C_i\right)^{-1};$$
$$x_i = P_i \left(M_i^{-1} A x_{i_f}[k-1] + C_i^T R_i^{-1} y_i[k]\right);$$

Step 2: local fusion

for each received $p_j(x[k])$, do

$\quad x_j = x_j[k], \quad P_j = P_j[k];$
$\quad \Gamma = \text{MutualCovariance}(P_i, P_j): \qquad (10) \wedge (11);$
$\quad \gamma = \text{MutualMean}(P_i, P_j, \Gamma, x_i, x_j): \qquad (14) \wedge (15);$
$\quad x_i = \left(P_i^{-1} + P_j^{-1} - \Gamma^{-1}\right)^{-1} \left(P_i^{-1} x_i + P_j^{-1} x_j - \Gamma^{-1} \gamma\right);$
$\quad P_i = \left(P_i^{-1} + P_j^{-1} - \Gamma^{-1}\right)^{-1};$

end

$$x_{i_f}[k] = x_i, \quad P_{i_f}[k] = P_i. \qquad \qquad \square$$

The second state-estimator performs the same operations as the DEI, with the difference that the CI [7] method is employed in the local fusion step instead of EI. As such, the second estimator is denoted with decentralized Covariance Intersection (DCI). In contrast to sending states, the third state-estimator sends measurements, i.e., y_i and R_i, which are then processed in a local Kalman filter (LKF) [3]. For all estimators let us define that $x_i[0] = (300, \cdots, 300)^\top$, $P_i[0] = 10I$, $Q = 1000I$ and $R_i = 1$. The results of the DEI, DCI and LKF, for node 3 and 5 are presented in Figure 8.
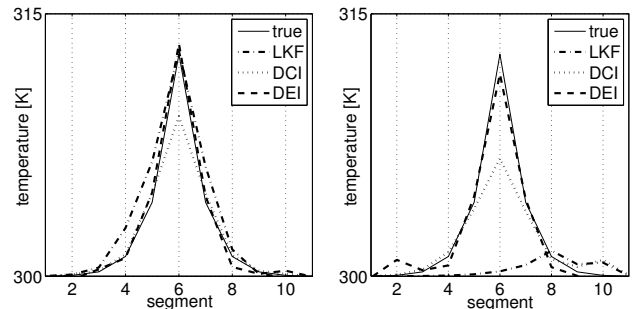
Fig. 8. Results of the estimated state-vector at node 3 (left) and node 5 (right) according to the DEI, DCI and LKF.

Figure 8 shows that local state-estimation based on local measurements limits the information that is available in the network. This is mainly noticed in the estimate of node 5, to which only information about the temperature in segment 8 and 10 is available. Node 5 of the DCI has an improved estimate compared to the LKF. However, due to

the fact that only $\mathrm{tr}(P_i)$ and $\mathrm{tr}(P_j)$ are used for state-fusion, uncertain estimates are weighted equally as accurate ones. As a result, the temperature distributions of the nodes 3 and 5 are somewhat averaged. The DEI on the other has a good estimate of the temperature distribution in both nodes. Moreover, whereas the estimates of the LKF and DCI in node 3 and 5 differ, the DEI also performs consensus on the state-estimates of the different nodes. Therefore the DEI outperforms both the LKF as well as the DCI.

VI. Conclusions

Current fusion methods of two estimates tend to bypass the mutual information and directly optimize the fused estimate. In contrast to that, this paper proposed a novel fusion method in which the mutual information results in an additional estimate, which defines a mutual mean and covariance. Both variables are derived from the two initial estimates. The mutual covariance was used to optimize the fused covariance, while the mutual mean was employed to optimize the fused mean. An example of decentralized state estimation, where the proposed fusion method is applied, showed a reduction in estimation error compared to the existing state fusion alternative algorithms.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transaction of the ASME Journal of Basic Engineering*, vol. 82, no. D, pp. 35–42, 1960.

[2] J. L. Speyer, "Computation and transmission requirements for a decentralized Linear-Quadratic-Gaussian control problem," *IEEE Transactions on Automatic Control*, vol. 24, no. 2, pp. 266–269, 1979.

[3] H. F. Durant-Whyte, B. Y. S. Rao, and H. Hu, "Towards a fully decentralized architecture for multi-sensor data fusion," in *1990 IEEE Int. Conf. on Robotics and Automation*, Cincinnati, Ohio, USA, 1990, pp. 1331–1336.

[4] R. Saber-Olfati, "Distributed Kalman filters for sensor networks," in *46th IEEE Conf. on Decision and Control*, New Orleans, LA, USA, 2007.

[5] A. Speranzon, C. Fischione, K. Johansson, and A. Sangiovanni-Vincentelli, "A Distributed Minimum Variance Estimator for Sensor Networks," *IEEE journal in selected areas in communications*, vol. 26, 2008.

[6] S. Julier, "Estimating and Exploiting the Degree of Independent Information in Distributed Data Fusion," in *In Proceedings of the 12th International Conference on Information Fusion*, 2009, pp. 772–779.

[7] S. J. Julier and J. K. Uhlmann, "A Non-divergent Estimation Algorithm in the Presence of Uknown Correlations," in *In Proc. of the American Control Conference*, Piscataway, NJ, USA, 1997, pp. 2369–2373.

[8] D. Franken and A. Hupper, "Improved Fast Covariance Intersection for Distributed Data Fusion," in *In Proc. of the 8th Int. Conf. on Information Fusion*, Philidalphia, PA, USA, 2005.

[9] L. Chen, P. Arambel, and R. Mehra, "Fusion under Unknown Correlation - Covariance Intersection as a Special Case," in *In Proceedings of 5th IEEE Int. Conf. on Information Fusion*, 2002, pp. 905–912.

[10] Y. Zhuo and J. Li, "Data Fusion of Unknown Correlations using Internal Ellipsoidal Approximations," in *Proceedings of the 17th IFAC World Congress*, 2008, pp. 2856–2860.

[11] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Mutlivariate analysis*. Academic press, London, 1979.

[12] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle filter for tracking applications*. Boston, Massachusetts: Artech House, 2004.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[14] D. S. Bernstein, *Matrix Mathematics*. Princeton University Press, 2005.

Appendix

A. Proof of Theorem IV.4

Let us first prove that $\varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x) \subseteq \varepsilon_{0,\Upsilon}(x)$ implies $x^\top \Upsilon^{-1} x \le 1$, for all $x \in \varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x)$, i.e., implies (9b). Let $\mathscr{Y} := \{\Upsilon \in \mathbb{R}^{n \times n} | \varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x) \subseteq \varepsilon_{0,\Upsilon}(x)\}$. Given a $x \in \varepsilon_{0,P_i}(x) \cup \varepsilon_{0,P_j}(x)$, then it holds that $x \in \varepsilon_{0,\Upsilon}(x)$, for all $\Upsilon \in \mathscr{Y}$ and thus, $x^\top \Upsilon^{-1} x \le 1$ for all $\Upsilon \in \mathscr{Y}$.

The next step is to show that $\Gamma_{\max} = \arg\min_\Upsilon \sum_q \lambda_q(\Upsilon)$. As $\lambda_q(\Upsilon) > 0$ for all $q \in \mathbb{Z}_{[1,n]}$, it holds that $\arg\min_\Upsilon \sum_q \lambda_q(\Upsilon) = \arg\min_\Upsilon \sum_q \log(\lambda_q(\Upsilon))$. Then, applying Fact 5.11.28 of [14], i.e., for any $\Upsilon \succ 0$ it holds that $|\Upsilon| = \prod_q \lambda_q(\Upsilon)$, one can derive that $\sum_q \log(\lambda_q(\Upsilon)) = \log\left(\prod_q \lambda_q(\Upsilon)\right) = \log|\Upsilon|$, which completes the proof. ∎

B. Proof of Lemma IV.5

Let $\hat{\Gamma} = \arg\min_{\hat{\Upsilon}} \log|\hat{\Upsilon}|$ and $\Gamma = \arg\min_\Upsilon \log|\Upsilon|$, where $\hat{\Upsilon}, \Upsilon \in \mathbb{R}^{n \times n}$. As $\Upsilon = S^{-1} D^{-1} \hat{\Upsilon} D^{-1} S$, the claim is proven if: (i) $\arg\min_{\hat{\Upsilon}} \log|\hat{\Upsilon}| = \arg\min_{\hat{\Upsilon}} \log|\Upsilon|$ and (ii) $\hat{x} \hat{\Upsilon}^{-1} \hat{x} \le 1$ for all $\hat{x} \in \varepsilon_{0,\hat{A}}(\hat{x}) \cup \varepsilon_{0,\hat{B}}(\hat{x})$, implies $x^T \Upsilon^{-1} x \le 1$ for all $x \in \varepsilon_{0,A}(x) \cup \varepsilon_{0,B}(x)$.

Let us start with (ii). By applying Proposition 2.6.9 of [14], i.e., $(EF)^{-1} = F^{-1} E^{-1}$ for any invertible matrices E and F, and the fact that $S = S^{-\top}$ one can derive that $x^\top A^{-1} x = x^\top (S^{-1} D^{-1} \hat{A} D^{-1} S)^{-1} x = x^\top S^{-1} D \hat{A}^{-1} D S x = x^\top S^\top D \hat{A}^{-1} D S x$. Hence, $x^\top A^{-1} x = \hat{x}^\top \hat{A}^{-1} \hat{x}$ and similarly $x^\top B^{-1} x = \hat{x}^\top \hat{B}^{-1} \hat{x}$ and $x^\top \Upsilon^{-1} x = \hat{x}^\top \hat{\Upsilon}^{-1} \hat{x}$. Therefore, if $\hat{x} \in \varepsilon_{0,\hat{A}}(\hat{x}) \cup \varepsilon_{0,\hat{B}}(\hat{x})$ then also $x \in \varepsilon_{0,A}(x) \cup \varepsilon_{0,B}(x)$ and if $\hat{x}^\top \hat{\Upsilon}^{-1} \hat{x} \le 1$ then also $x^\top \Upsilon^{-1} x \le 1$, which proves (ii).

The proof of (i) starts with $\log|\Upsilon| = \log|S^{-1} D^{-1} \hat{\Upsilon} D^{-1} S|$. Applying Proposition 2.7.3 and Corollary 2.7.4 of [14], i.e., $|EF| = |E||F|$ and $|E^{-1}| = |E|^{-1}$ holds for any nonsingular matrices E, F, gives that $\log|S^{-1} D^{-1} \hat{\Upsilon} D^{-1} S| = \log|\hat{\Upsilon}| + 2\log|D|^{-1}$. Hence, $\arg\min_{\hat{\Upsilon}} \log(\Upsilon) = \arg\min_{\hat{\Upsilon}} \log(\hat{\Upsilon}) + \log|D|^{-1} = \arg\min_{\hat{\Upsilon}} \log(\hat{\Upsilon})$, which completes the proof. ∎

C. Proof of Theorem IV.6

Let us define the transformation $\hat{x} := S_j^{-1} D_i^{-0.5} S_i^{-1} x$ and similarly $\hat{P}_i := S_j^{-1} D_i^{-0.5} S_i^{-1} P_i S_i D_i^{-0.5} S_j$ and $\hat{P}_j := S_j^{-1} D_i^{-0.5} S_i^{-1} P_j S_i D_i^{-0.5} S_j$. Notice that each column of S_i is an eigenvector of P_i. As P_i is a symmetric matrix, Corollary 5.4.8 of [14] gives that all its eigenvectors are orthogonal to each other, i.e., $S_i^\top S_i = I$ and $S_i = S_i^{-\top}$. Similarly, $S_j = S_j^{-\top}$ also holds. Hence, we can apply Lemma IV.5 with $D_\Gamma = S_j^{-1} D_i^{-0.5} S_i^{-1} \Gamma S_i D_i^{-0.5} S_j$. The last step of the proof is to show that $\varepsilon_{0,D_\Gamma}(\hat{x})$ is the LJE of $\varepsilon_{0,\hat{P}_i}(\hat{x}) \cup \varepsilon_{0,\hat{P}_j}(\hat{x})$.

From (10) it follows that $\hat{P}_i = I$ and $\hat{P}_j = D_j$ are both diagonal matrices. Hence, let us search for the LJE $\varepsilon_{0,D_\Gamma}(\hat{x})$ such that D_Γ is a diagonal matrix as well. The condition of this LJE, i.e., $\varepsilon_{0,I}(\hat{x}) \cup \varepsilon_{0,D_j}(\hat{x}) \subseteq \varepsilon_{0,D_\Gamma}(\hat{x})$, originates from the inequalities $D_\Gamma \succeq I$ and $D_\Gamma \succeq D_j$ and thus:

$$\lambda_q(D_\Gamma) \ge 1 \quad \text{and} \quad \lambda_q(D_\Gamma) \ge \lambda_q(D_j), \quad \forall q \in \mathbb{Z}_{[1,n]}. \quad (16)$$

Therefore, minimization of $\sum_q \lambda_q(D_\Gamma)$ while satisfying (16) implies that $\lambda_q(D_\Gamma) := \max(\lambda_q(D_j), 1)$, for all $q \in \mathbb{Z}_{[1,n]}$. As D_j and D_Γ are diagonal matrices it follows that $\lambda_q(D_\Gamma) = [D_\Gamma]_{qq}$ and $\lambda_q(D_j) = [D_j]_{qq}$, which completes the proof. ∎