A Quarter Century of Covariance Intersection: Correlations Still Unknown? [Lecture Notes]

Robin Forsling, Benjamin Noack and Gustaf Hendeby

Over the past two and a half decades, covariance intersection (CI) has provided means for robust estimation in scenarios where the uncertainty information is incomplete. Estimation in distributed and decentralized data fusion settings is typically characterized by having non-zero cross-correlations between the estimates to be merged. Mean square error (MSE) optimal estimators such as the Kalman filter (KF) are limited to data fusion where these cross-correlations are fully known. Keeping track of cross-correlations is unfortunately not always possible. To quantify confidence in the estimate's uncertainty, the concept of *conservativeness* has been introduced. A conservative estimator guarantees that the computed covariance matrix is not smaller than the actual covariance matrix. It turns out that CI guarantees conservativeness for any degree of unknown cross-correlations as long as the estimates to be fused are conservative. It should be noted that, in the CI literature, the notion of *covariance consistency* is often used to characterize *conservativeness*. In this work, we use the latter term.

Estimation under unknown correlations is a common issue in decentralized data fusion (DDF). Modern application areas include air surveillance, autonomous vehicles, and vehicle-to-everything (V2X, [1]) networks. A V2X scenario is illustrated in Figure 1, where multiple autonomous vehicles use local sensors to measure other agents' states of the scene, for example, the positions of other vehicles and pedestrians. Measurements are preprocessed into local track estimates. The estimates are exchanged and, after that, fused in other agents. There is also broadcasting of information from pedestrians and ground stations, and datalink communication between infrastructure, satellites, and cloud services. The essence of Figure 1 is the vast amount of data that is required to be exchanged, which means that it quickly becomes impossible to keep track of which sensor has actually extracted what information. More technically, preprocessing and communication of estimates and information give rise to correlations, and these correlations are generally impossible to maintain knowledge about. If the correlations are neglected, fusion of a local estimate with another estimate using, for instance, a naïve fusion rule produces an estimate with an incorrect uncertainty such that the error covariance matrix does not conservatively bound the actual error. This issue is circumvented if instead CI is used.

CI has proven to be relevant in numerous and diverse application areas. A typical example is the object tracking setting, for instance, in [2], but it does not end there. Simultaneous localization and mapping (SLAM) applications using CI are studied in [3]–[6]. CI has further been used for problems ranging from localization of multi-robot systems [7] and cooperative localization of robot swarms [8] over time-of-arrival localization [9] to automotive tracking, human tracking [10], data validation [11], and medical application [12]. Below we shall see what unites the applications mentioned above and what awaits CI in the future. An overview of estimation techniques applicable for distributed and decentralized estimation is provided in *Sidebar: A Brief Survey of Distributed and Decentralized Estimation.*

Conservativeness

We start by formally defining conservativeness. Let $\mathbf{x}$ be an unknown state to be estimated. An estimate of $\mathbf{x}$ is given by $\hat{\mathbf{x}}$ and an error covariance matrix $\mathbf{P}$, where $\mathbf{P}$ captures the uncertainty of $\hat{\mathbf{x}}$. The estimate $\hat{\mathbf{x}}$ is conservative if

$$\mathbf{P} \succeq \mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}] \ , \tag{1}$$

where $\mathbb{E}$ is the expectation operator, $\tilde{\mathbf{x}} = \hat{\mathbf{x}} - \mathbf{x}$ is the estimation error and $\mathbf{A} \succeq \mathbf{B}$ means the difference $\mathbf{A} - \mathbf{B}$ is positive semidefinite. The special case $\mathbf{P} = \mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}]$ means that the estimate's uncertainty is exactly known.

To illustrate the relationship, let $\mathcal{E}(\mathbf{S}) = \{\mathbf{z} \in \mathbb{R}^n \,|\, \mathbf{z}^\mathsf{T}\mathbf{S}^{-1}\mathbf{z} \leq 1\}$ denote the ellipsoid of an $n \times n$ positive definite matrix $\mathbf{S}$ centered at the origin. A geometric interpretation of (1) can be seen in Figure 2, where two ellipses $\mathcal{E}(\mathbf{P})$ and $\mathcal{E}(\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}])$ are illustrated. The relation $\mathbf{P} \succeq \mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}]$ is equivalent to $\mathcal{E}(\mathbf{P}) \supseteq \mathcal{E}(\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}])$.

Optimal and Conservative Fusion

To illustrate the data fusion problem, assume that $\mathbf{x}$ is a 2D state vector. Two state estimates $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ are available, which are related to the state $\mathbf{x}$ through

$$\begin{aligned} \tilde{\mathbf{x}}^1 &= \mathbf{x} + \hat{\mathbf{x}}^1, \qquad & \mathbf{P}^1 &= \mathbb{E}[\tilde{\mathbf{x}}^1(\tilde{\mathbf{x}}^1)^\mathsf{T}], \\ \tilde{\mathbf{x}}^2 &= \mathbf{x} + \hat{\mathbf{x}}^2, \qquad & \mathbf{P}^2 &= \mathbb{E}[\tilde{\mathbf{x}}^2(\tilde{\mathbf{x}}^2)^\mathsf{T}], \end{aligned}$$

where $\tilde{\mathbf{x}}^i = \hat{\mathbf{x}}^i - \mathbf{x}$ is the error and $\mathbf{P}^i$ is the error covariance matrix of the ith estimate. The cross-covariance matrix $\mathbf{P}^{12} = \mathbb{E}[\tilde{\mathbf{x}}^1(\tilde{\mathbf{x}}^2)^\mathsf{T}]$ describes the correlations between $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$. Fusion aims at determining the gains $\mathbf{K}$ and $\mathbf{L} = \mathbf{I} - \mathbf{K}$ such that the combination

$$\hat{\mathbf{x}}^{\text{fus}} = \mathbf{K}\,\hat{\mathbf{x}}^1 + \mathbf{L}\,\hat{\mathbf{x}}^2$$

results in an optimal error covariance matrix

$$\begin{aligned} \mathbf{P}^{\text{fus}} &= \mathbb{E}[\tilde{\mathbf{x}}^{\text{fus}}(\tilde{\mathbf{x}}^{\text{fus}})^\mathsf{T}] \\ &= \mathbf{K}\mathbf{P}^1\mathbf{K}^\mathsf{T} + \mathbf{K}\mathbf{P}^{12}\mathbf{L}^\mathsf{T} + \mathbf{L}(\mathbf{P}^{12})^\mathsf{T}\mathbf{K}^\mathsf{T} + \mathbf{L}\mathbf{P}^2\mathbf{L}^\mathsf{T} \ , \end{aligned} \tag{2}$$

where

$$\tilde{\mathbf{x}}^{\text{fus}} = \hat{\mathbf{x}}^{\text{fus}} - \mathbf{x}^{\text{fus}} = \mathbf{K}\,\tilde{\mathbf{x}}^1 + \mathbf{L}\,\tilde{\mathbf{x}}^2$$

is the estimation error, and optimality typically refers to minimizing the mean square estimation (MSE) error given by $\text{tr}(\mathbf{P}^{\text{fus}})$. If the cross-covariance matrix is known, an MSE optimal fusion result is attained using the Bar-Shalom–Campo formulas [13] with the specific gains

$$\mathbf{K}^{\text{BC}} = (\mathbf{P}^2 - (\mathbf{P}^{12})^\mathsf{T})\left(\mathbf{P}^1 + \mathbf{P}^2 - \mathbf{P}^{12} - (\mathbf{P}^{12})^\mathsf{T}\right)^{-1} , \tag{3a}$$

$$\mathbf{L}^{\text{BC}} = (\mathbf{P}^1 - \mathbf{P}^{12})\left(\mathbf{P}^1 + \mathbf{P}^2 - \mathbf{P}^{12} - (\mathbf{P}^{12})^\mathsf{T}\right)^{-1} . \tag{3b}$$

The fused estimate then becomes

$$\hat{\mathbf{x}}^{\text{BC}} = \mathbf{K}^{\text{BC}}\hat{\mathbf{x}}^1 + \mathbf{L}^{\text{BC}}\hat{\mathbf{x}}^2 = \hat{\mathbf{x}}^1 + \mathbf{L}^{\text{BC}}\left(\hat{\mathbf{x}}^2 - \hat{\mathbf{x}}^1\right) , \tag{4a}$$

$$\begin{aligned} \mathbf{P}^{\text{BC}} &= \mathbf{K}^{\text{BC}}\mathbf{P}^1(\mathbf{K}^{\text{BC}})^\mathsf{T} + \mathbf{K}^{\text{BC}}\mathbf{P}^{12}(\mathbf{L}^{\text{BC}})^\mathsf{T} \\ &\quad + \mathbf{L}^{\text{BC}}(\mathbf{P}^{12})^\mathsf{T}(\mathbf{K}^{\text{BC}})^\mathsf{T} + \mathbf{L}^{\text{BC}}\mathbf{P}^2(\mathbf{L}^{\text{BC}})^\mathsf{T} \\ &= \mathbf{P}^1 - \mathbf{L}^{\text{BC}}\left(\mathbf{P}^1 + \mathbf{P}^2 - \mathbf{P}^{12} - (\mathbf{P}^{12})^\mathsf{T}\right)^{-1}(\mathbf{L}^{\text{BC}})^\mathsf{T} \ , \end{aligned} \tag{4b}$$

which also has the smallest covariance matrix $\mathbf{P}^{\text{BC}} \preceq \mathbf{P}^{\text{fus}}$ of all possible fusion results (2). A consequence of this property is that $\mathbf{P}^{\text{BC}}$

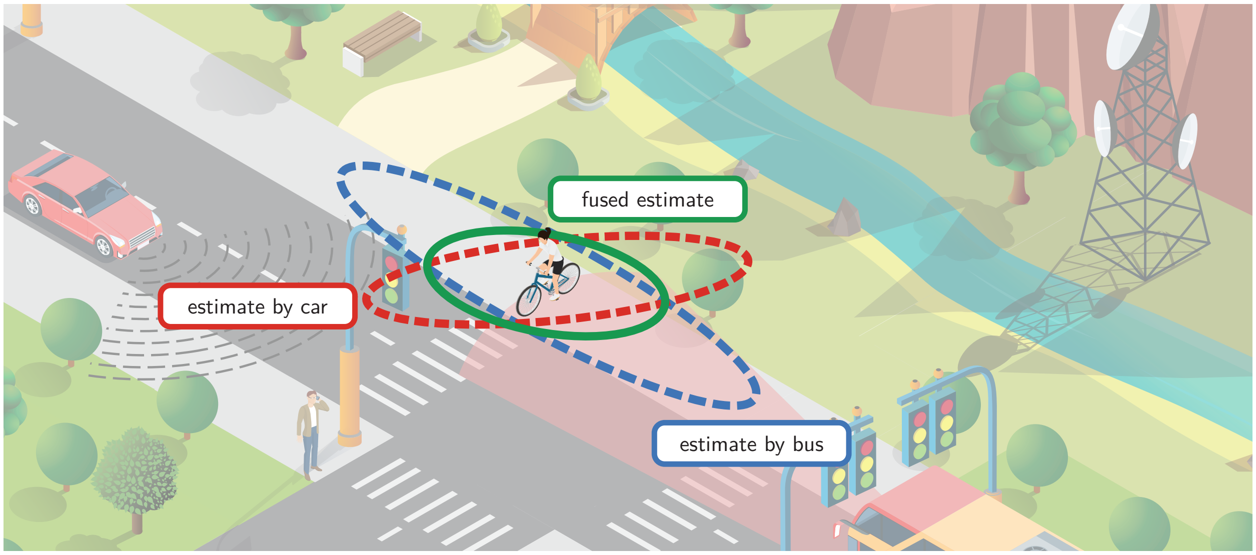


Figure 1: In a vehicle-to-everything network as shown on the front cover, multiple vehicles measure the surroundings and preprocess the measurement information locally for situational awareness. Local estimates are communicated to, for instance, other vehicles, ground stations, and cloud services. Information flows in arbitrary directions and cross-correlations between local estimates are hard to keep track of. With CI, information from different sources can be fused reliably.

and, thus, $\mathbf{K}^{\mathrm{BC}}$ and $\mathbf{L}^{\mathrm{BC}}$ are optimal for the trace, determinant, and all other cost functions J that have the property $\mathbf{P}^{\mathrm{A}} \preceq \mathbf{P}^{\mathrm{B}} \Rightarrow J(\mathbf{P}^{\mathrm{A}}) \leq J(\mathbf{P}^{\mathrm{B}})$. Each possible cross-covariance matrix $\mathbf{P}^{12}$ yields a different result $\mathbf{P}^{\mathrm{BC}}$, where each $\mathcal{E}(\mathbf{P}^{\mathrm{BC}})$ lies in the intersection of $\mathcal{E}(\mathbf{P}^1)$ and $\mathcal{E}(\mathbf{P}^2)$. Figure 3 illustrates this relationship for two 2D estimates and different admissible $\mathbf{P}^{12}$.

Many DDF systems encounter the problem of an incorrectly determined cross-covariance matrix $\mathbf{P}^{12}$ or are even ignorant about it. The task is then to compute a fused estimate $\hat{\mathbf{x}}$ and its covariance matrix $\mathbf{P}$ from $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ given that $\mathbf{P}^{12}$ is unknown. CI as proposed by [14] preserves conservativeness irrespective of any underlying correlations and fuses two estimates by

$$\hat{\mathbf{x}}^{\mathrm{CI}} = \mathbf{P}^{\mathrm{CI}}\Big(\omega(\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + (1-\omega)(\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2\Big) , \tag{5a}$$

$$\mathbf{P}^{\mathrm{CI}} = \Big(\omega(\mathbf{P}^1)^{-1} + (1-\omega)(\mathbf{P}^2)^{-1}\Big)^{-1} , \tag{5b}$$

where $\omega \in [0,1]$ is found by optimizing $J(\mathbf{P}^{\mathrm{CI}})$. In contrast to the Bar-Shalom–Campo fusion, the optimal CI estimate is not unique for different criteria J. For instance, minimizing the trace or determinant leads to different estimates (5). The results of fusing $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ are visualized in Figure 3 revealing the geometrical interpretation of CI: $\mathcal{E}(\mathbf{P}^{\mathrm{CI}})$ encloses the intersection of $\mathcal{E}(\mathbf{P}^1)$ and $\mathcal{E}(\mathbf{P}^2)$, which gives it the name *covariance intersection.* To be clear, the actual (yet unknown) error ellipse $\mathcal{E}(\mathbb{E}[(\hat{\mathbf{x}}^{\mathrm{CI}} - \mathbf{x})(\hat{\mathbf{x}}^{\mathrm{CI}} - \mathbf{x})^{\mathsf{T}}])$ of $\hat{\mathbf{x}}^{\mathrm{CI}}$ does not need to be inside the intersection.

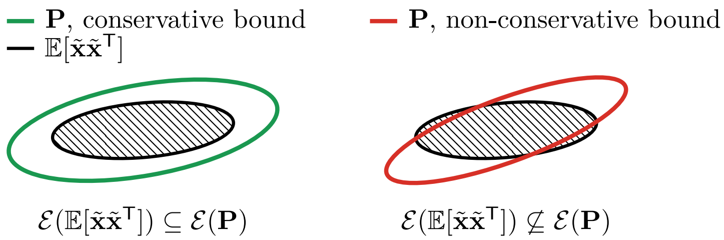


Figure 2: Geometric interpretation of conservativeness. The colored ellipses represent the covariance matrix computed by an estimator, and the hatched area represents the covariance matrix of the actual error. *Left:* Conservative bound where $\mathcal{E}(\mathbf{P}) \supseteq \mathcal{E}(\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}])$. *Right:* Non-conservative bound for which $\mathcal{E}(\mathbf{P}) \not\supseteq \mathcal{E}(\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}])$.

However, it is conservatively bounded by $\mathcal{E}(\mathbf{P}^{\mathrm{CI}})$, that is, the green ellipse in Figure 3, meaning that CI is conservative for all admissible values of $\mathbf{P}^{12}$ and hence

$$\mathbf{P}^{\mathrm{CI}} \succeq \mathbb{E}[(\hat{\mathbf{x}}^{\mathrm{CI}} - \mathbf{x})(\hat{\mathbf{x}}^{\mathrm{CI}} - \mathbf{x})^\mathsf{T}]$$

is guaranteed.

This and other critical properties of CI are discussed in the first part of this article. We will identify several challenges in applying CI and provide an overview of current approaches and solutions, and try to find answers to the question: *Are correlations still unknown?*

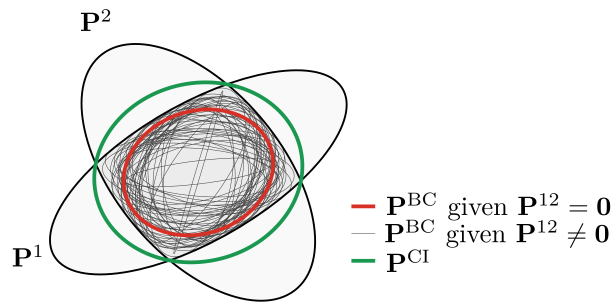


Figure 3: Geometric interpretation of covariance intersection. The example illustrates fusion of two estimates with covariance matrices $\mathbf{P}^1$ and $\mathbf{P}^2$. Multiple $\mathbf{P}^{\mathrm{BC}}$ are generated based on different assumptions on $\mathbf{P}^{12}$, including the uncorrelated case $\mathbf{P}^{12} = \mathbf{0}$. The covariance matrix of a conservative estimate is given by $\mathbf{P}^{\mathrm{CI}}$ where the ellipse of $\mathbf{P}^{\mathrm{CI}}$ encloses the intersection of $\mathbf{P}^1$ and $\mathbf{P}^2$, including the ellipses of all possible $\mathbf{P}^{\mathrm{BC}}$. Here, CI uses the trace to determine ω.

Covariance Intersection

The primary benefit of centralized data fusion over DDF is that optimal estimation is possible. Hence, it is possible to design the data fusion network to allow for optimal information utilization. However, centralized data fusion exhibits a single critical point of failure. Critical functionality is not distributed and therefore the complexity of large-scale systems quickly explodes. Network topologies and their properties are further discussed in, for example, [15]–[18].

With the applications mentioned at the beginning, distributed and even fully decentralized data fusion is gaining increasing importance. DDF scenarios in which no correlations are present and CI is irrelevant are rare. The reason is that CI conservatively handles any degree of cross-correlations, and cross-correlations are an intrinsic part of almost every DDF problem. The motivation for developing DDF networks even in the presence of such dependencies is: (i) *Robustness*. There is no single critical point of failure, the DDF network is fault-tolerant. (ii) *Modularity*. Large-scale systems can be decomposed into smaller modules, which are self-contained and, hence, reduce the system complexity. (iii) *Flexibility*. Addition, modification, and removal of agents are possible on the fly. These are major advantages over centralized structure and motivate the study and use of CI.

Problem Intuition

To get an intuition for how CI works and performs in different situations, we use a simple case where $\mathbf{x}$ is 2D. To begin with, consider that two estimates $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ of $\mathbf{x}$ are available with covariance matrices $\mathbf{P}^1$ and $\mathbf{P}^2$ as defined by their ellipses in Figure 4, which are of the same shape but have different orientations. It is assumed that $\mathbf{P}^{12} = \mathbf{0}$. In the left part, the ellipses of $\mathbf{P}^1$ and $\mathbf{P}^2$ are almost aligned, and to the right they are perpendicular. Merging the two estimates using CI optimized on trace yields $\mathbf{P}^{\text{CI}}$. Also, $\mathbf{P}^{\text{BC}}$ given $\mathbf{P}^{12} = \mathbf{0}$ is computed. As it can be seen, if $\mathbf{P}^1 \approx \mathbf{P}^2$ then $\text{tr}(\mathbf{P}^{\text{CI}}) \approx \text{tr}(\mathbf{P}^1) = \text{tr}(\mathbf{P}^2)$ and there is no significant gain from merging $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$. On the other hand, for the geometry in the right part, the information gain is significant, and $\text{tr}(\mathbf{P}^{\text{CI}}) < \text{tr}(\mathbf{P}^1) = \text{tr}(\mathbf{P}^2)$. However, $\mathbf{P}^{\text{CI}} = 2\mathbf{P}^{\text{BC}}$ as CI respects any possible cross-covariance matrix $\mathbf{P}^{12}$.

This example demonstrates how the alignment of the uncertainties affects the performance of CI as compared to optimal fusion with BC. The discrepancy between CI and BC is reduced when the intersection area is small, as in the right part of Figure 4. Hence, CI typically performs well when error covariances have different orientations: The errors of an estimate $\hat{\mathbf{x}}^1$ are small in directions where estimate $\hat{\mathbf{x}}^2$ has large errors and vice versa. The worst performance is seen for $\mathbf{P}^1 \approx \mathbf{P}^2$, and when $\mathbf{P}^1 \preceq \mathbf{P}^2$ or $\mathbf{P}^1 \succeq \mathbf{P}^2$. In the latter two cases, CI will simply select $\hat{\mathbf{x}}^1$ or $\hat{\mathbf{x}}^2$, respectively, meaning that either $\omega = 1$ or $\omega = 0$ is computed in (5). CI has to account for the possibility of fully correlated estimates, which justifies this result. In these cases, one can choose other cost functions J than the trace to find an ω that lies between 0 and 1.

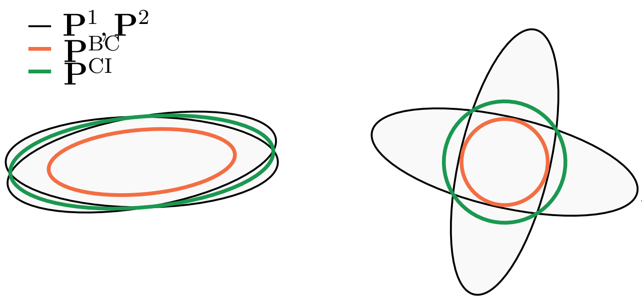


Figure 4: Impact of the geometry on the fusion result when using covariance intersection. *Left:* $\mathbf{P}^1 \approx \mathbf{P}^2$ and the fusion gain is small, that is, $\mathrm{tr}(\mathbf{P}^{\mathrm{CI}}) \approx \mathrm{tr}(\mathbf{P}^1)$. *Right*: If $\mathbf{P}^2$ is equal to $\mathbf{P}^1$ rotated by 90° the fusion gain is increased such that $\mathrm{tr}(\mathbf{P}^1) \gg \mathrm{tr}(\mathbf{P}^{\mathrm{CI}})$.

Mathematical Background

In data fusion, the information form [19] is a convenient way of representing an estimate $\hat{\mathbf{x}}$ with covariance matrix $\mathbf{P}$. The information form uses the information vector $\mathbf{P}^{-1}\hat{\mathbf{x}}$ and the information matrix $\mathbf{P}^{-1}$. For example, consider an MSE optimal estimator with $\mathbf{P}^{12} = \mathbf{0}$. In this case, (3a) and (3b) become

$$\mathbf{K} = \mathbf{P}^2(\mathbf{P}^1 + \mathbf{P}^2)^{-1} = \left((\mathbf{P}^1 + \mathbf{P}^2)(\mathbf{P}^2)^{-1}\right)^{-1} = \left(\mathbf{P}^1(\mathbf{P}^2)^{-1} + \mathbf{I}\right)^{-1} = \left((\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1}\right)^{-1}(\mathbf{P}^1)^{-1}$$

and $\mathbf{L} = ((\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1})^{-1}(\mathbf{P}^2)^{-1}$, respectively. Hence, the fusion result (4) reduces to

$$\hat{\mathbf{x}} = \mathbf{P}(\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + \mathbf{P}(\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2, \quad (6a)$$
$$\mathbf{P} = \left((\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1}\right)^{-1}, \quad (6b)$$

which becomes

$$\mathbf{P}^{-1}\hat{\mathbf{x}} = (\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + (\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2 , \quad (7a)$$
$$\mathbf{P}^{-1} = (\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1} \quad (7b)$$

in the information form. The result is optimal for zero cross-correlations. If these fusion formulas are erroneously used in the presence of non-zero cross-correlations $\mathbf{P}^{12} \neq \mathbf{0}$, fusion method (6) is typically referred to as *naïve fusion*.

As provided in [14], the CI formulas in (5) have the corresponding information form

$$(\mathbf{P}^{\mathrm{CI}})^{-1}\hat{\mathbf{x}}^{\mathrm{CI}} = \omega(\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + (1-\omega)(\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2, \quad (8a)$$
$$(\mathbf{P}^{\mathrm{CI}})^{-1} = \omega(\mathbf{P}^1)^{-1} + (1-\omega)(\mathbf{P}^2)^{-1} \quad (8b)$$

with $\omega \in [0,1]$. The resemblance between the naïve fusion rule (6) and CI is striking. The formulas are equivalent except for the parameter ω. The constraint $\omega \in [0,1]$ means that $(\mathbf{P}^{\mathrm{CI}})^{-1}$ is a convex combination of $(\mathbf{P}^1)^{-1}$ and $(\mathbf{P}^2)^{-1}$. It should noted that while [14] is the original paper suggesting CI, the basic ideas behind the method are actually developed already in [20] under the name Gaussian intersection.

The solution in (5), and equivalently in (8), is parametrized by $\omega \in [0,1]$. As a result, a family of solutions to the fusion problem is produced [21], as shown in Figure 5. To determine a particular value for ω, a loss function $J(\mathbf{P}^{\mathrm{CI}})$ is minimized. Common choices of J are the trace $\mathrm{tr}(\cdot)$ and the determinant $\det(\cdot)$, where $\mathrm{tr}(\mathbf{P}^{\mathrm{CI}})$ is related to the MSE of $\hat{\mathbf{x}}^{\mathrm{CI}}$ and $\det(\mathbf{P}^{\mathrm{CI}})$ is related to entropy [22]. Strictly speaking, finding the optimal value of ω is equivalent to solving

$$\underset{\omega\in[0,1]}{\text{minimize}} \quad J\left(\left(\omega(\mathbf{P}^1)^{-1} + (1-\omega)(\mathbf{P}^2)^{-1}\right)^{-1}\right). \quad (9)$$

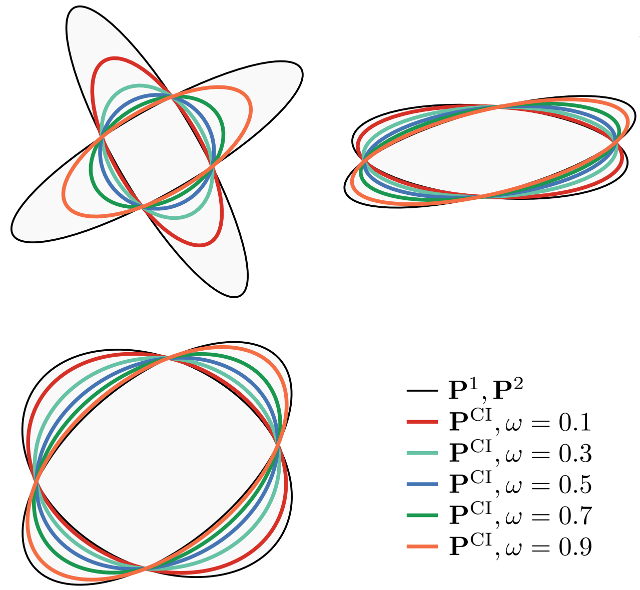


Figure 5: Visualization of how $\mathbf{P}^{\mathrm{CI}}$ depends on the weight parameter ω. Three cases with different $\mathbf{P}^1$ and $\mathbf{P}^2$ are illustrated. For each case, covariance intersection yields a family of solutions where each solution corresponds to a particular value of ω.

In the special case of merging 1D estimates, choosing ω is trivial. For example, assuming $\mathbf{P}^1 = 1$ and $\mathbf{P}^2 = 4$ yields

$$\mathbf{P}^{\mathrm{CI}} = \frac{1}{\omega + (1-\omega)\frac{1}{4}} \ ,$$

which is monotonic in $\omega \in [0, 1]$. In this case, $\mathrm{tr}(\mathbf{P}^{\mathrm{CI}})$ is minimized for $\omega = 1$ which means that $\hat{\mathbf{x}}^{\mathrm{CI}} = \hat{\mathbf{x}}^1$ and $\mathbf{P}^{\mathrm{CI}} = \mathbf{P}^1$. No actual fusion is hence performed.

Another important tool to study correlations and model fusion problems is a joint state-space representation. Two estimates $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ are expressed as the joint state $\hat{\mathbf{x}}^J = \begin{bmatrix} (\hat{\mathbf{x}}^1)^\mathsf{T} & (\hat{\mathbf{x}}^2)^\mathsf{T} \end{bmatrix}^\mathsf{T}$ with the corresponding joint covariance matrix

$$\mathbf{P}^J = \mathbb{E}[\tilde{\mathbf{x}}^J (\tilde{\mathbf{x}}^J)^\mathsf{T}] = \begin{bmatrix} \mathbf{P}^1 & \mathbf{P}^{12} \\ (\mathbf{P}^{12})^\mathsf{T} & \mathbf{P}^2 \end{bmatrix} \succeq \mathbf{0} \ , \tag{10}$$

where $\tilde{\mathbf{x}}^J = \begin{bmatrix} (\tilde{\mathbf{x}}^1)^\mathsf{T} & (\tilde{\mathbf{x}}^2)^\mathsf{T} \end{bmatrix}^\mathsf{T}$ is the joint error. This form will be used frequently in the subsequent sections. The optimal fusion rule (4) can now be formulated as a weighted least-squares problem (WLS) [23]

$$\begin{bmatrix} \hat{\mathbf{x}}^1 \\ \hat{\mathbf{x}}^2 \end{bmatrix} = \mathbf{H}^J \mathbf{x} + \tilde{\mathbf{x}}^J$$

with $\mathbf{H}^J = \begin{bmatrix} \mathbf{I} & \mathbf{I} \end{bmatrix}^\mathsf{T}$, and has the solution

$$(\mathbf{P}^{\mathrm{BC}})^{-1} \hat{\mathbf{x}}^{\mathrm{BC}} = (\mathbf{H}^J)^\mathsf{T} (\mathbf{P}^J)^{-1} \hat{\mathbf{x}}^J \ , \tag{11a}$$
$$(\mathbf{P}^{\mathrm{BC}})^{-1} = (\mathbf{H}^J)^\mathsf{T} (\mathbf{P}^J)^{-1} \mathbf{H}^J \ . \tag{11b}$$

This joint state-space formulation also exhibits the advantage that it can also express more complex relationships between estimates and state through $\mathbf{H}^J = \begin{bmatrix} (\mathbf{H}^1)^\mathsf{T} & (\mathbf{H}^2)^\mathsf{T} \end{bmatrix}^\mathsf{T}$, for example, in case of fusion of unequal state vectors [24]–[27].

The naïve fusion result is obtained when $\mathbf{P}^{12} = \mathbf{0}$ is used in (10) to ignore possible correlations. The joint covariance for CI is similarly given by

$$\mathbf{P}^{J,\mathrm{CI}} = \begin{bmatrix} \frac{\mathbf{P}^1}{\omega} & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{P}^2}{1-\omega} \end{bmatrix} , \tag{12}$$

which also sets the off-diagonal blocks to zero but inflates the diagonal blocks. As noted by, for instance, [21], [28], [29], these expressions result in (8), that is,

$$(\mathbf{P}^{\mathrm{CI}})^{-1} \hat{\mathbf{x}}^{\mathrm{CI}} = (\mathbf{H}^J)^\mathsf{T} (\mathbf{P}^{J,\mathrm{CI}})^{-1} \hat{\mathbf{x}}^J = \omega (\mathbf{P}^1)^{-1} \hat{\mathbf{x}}^1 + (1-\omega)(\mathbf{P}^2)^{-1} \hat{\mathbf{x}}^2 \ , \tag{13a}$$

$$(\mathbf{P}^{\mathrm{CI}})^{-1} = (\mathbf{H}^J)^\mathsf{T} (\mathbf{P}^{J,\mathrm{CI}})^{-1} \mathbf{H}^J = \omega (\mathbf{P}^1)^{-1} + (1-\omega)(\mathbf{P}^2)^{-1} \tag{13b}$$

We can use the joint state-space representation to conclude the mathematical background with a simple proof of CI's conservativeness. Completely unknown cross-correlations are mathematically equivalent to saying that $\mathbf{P}^{12}$ can be any matrix such that the joint covariance matrix (10) is positive semidefinite. Since

$$\mathbf{S} \succeq \mathbf{R} \implies (\mathbf{H}^J)^\mathsf{T} \mathbf{S} \mathbf{H}^J \succeq (\mathbf{H}^J)^\mathsf{T} \mathbf{R} \mathbf{H}^J \ , \tag{14}$$

it is sufficient to show $\mathbf{P}^{J,\mathrm{CI}} \succeq \mathbf{P}^J$ to conclude that CI is conservative. For $\omega \in (0,1)$, we first note that positive semidefiniteness of (10) is equivalent to $\mathbf{P}^1 \succeq \mathbf{0}$ and $\mathbf{P}^1 - \mathbf{P}^{12}(\mathbf{P}^2)^{-1}(\mathbf{P}^{12})^\mathsf{T} \succeq \mathbf{0}$, where the latter is the Schur complement. Hence,

$$\begin{aligned}\mathbf{P}^{J,\mathrm{CI}} - \mathbf{P}^J &= \begin{bmatrix} \frac{\mathbf{P}^1}{\omega} - \mathbf{P}^1 & -\mathbf{P}^{12} \\ -(\mathbf{P}^{12})^\mathsf{T} & \frac{\mathbf{P}^2}{1-\omega} - \mathbf{P}^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1-\omega}{\omega}\mathbf{P}^1 & -\mathbf{P}^{12} \\ -(\mathbf{P}^{12})^\mathsf{T} & \frac{\omega}{1-\omega}\mathbf{P}^2 \end{bmatrix} \succeq \mathbf{0}\end{aligned}$$

is equivalent to $\frac{1-\omega}{\omega}\mathbf{P}^1 \succeq \mathbf{0}$ and

$$\frac{1-\omega}{\omega}\mathbf{P}^1 - \frac{1-\omega}{\omega}\mathbf{P}^{12}(\mathbf{P}^2)^{-1}(\mathbf{P}^{12})^\mathsf{T} \succeq \ \mathbf{0}$$

$$\iff \mathbf{P}^1 - \mathbf{P}^{12}(\mathbf{P}^2)^{-1}(\mathbf{P}^{12})^\mathsf{T} \succeq \ \mathbf{0} \ ,$$

where it was used that $(1-\omega)/\omega > 0$. The last statement is true because (10) is positive semidefinite, and therefore CI is conservative for any possible $\mathbf{P}^{12}$. The cases $\omega \in \{0,1\}$ are handled trivially since in these cases there is no actual fusion. CI in the form (13) can be interpreted as covariance inflation with $\mathbf{P}^{J,\mathrm{CI}} \succeq \mathbf{P}^J$.

Relationship (14) has strong implications to the design and study of conservative fusion methods: A WLS-based fusion (11) becomes a conservative fusion method if $\mathbf{P}^J$ in (11) is replaced by an inflated version $\mathbf{P}^{\mathrm{infl}}$, such that $\mathbf{P}^{\mathrm{infl}} \succeq \mathbf{P}^J$ holds for any possible $\mathbf{P}^{12}$.

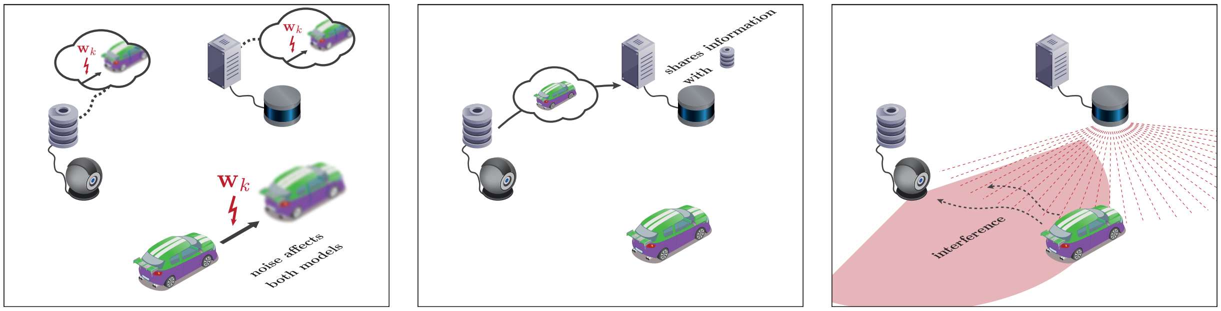
Disturbing Disturbances

Data fusion depends on dependencies, and dependencies are quantified by correlations between estimates. In the following, we look at how it is possible to have $\mathbf{P}^{12} \neq \mathbf{0}$ while initially having uncorrelated estimates. Three main sources of correlations can be identified: *common process noise*, *common information*, and *correlated sensor noise*. Identifying situations where cross-correlations arise is crucial to be able to properly account for the cross-correlations. Further discussions of how to identify and handle dependencies are, for instance, provided in [30], [31]. The three considered cases are illustrated in Figure 6.

Common Process Noise: Recursive formulas of cross-correlations are presented in [32]. Two estimates of the same state become correlated due to common process noise [33] even if otherwise independent agents compute them. At first glance, this might seem confusing, but consider two estimates $\hat{\mathbf{x}}^1_{k|k}$ and $\hat{\mathbf{x}}^2_{k|k}$ of a dynamic state $\mathbf{x}_k$, where k is a time index. Assume $\mathbf{P}^{12}_{k|k} = \mathbf{0}$ and that $\mathbf{x}_k$ evolves according to

$$\mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{w}_k \ ,$$

where $\mathbf{F}_k$ is the state transition model and $\mathbf{w}_k$ is a zero-mean process noise with $\mathbf{Q}_k = \mathbb{E}[\mathbf{w}_k\mathbf{w}_k^\mathsf{T}]$. The predicted state estimates in the Kalman filters used by the two agents



((a)) Common process noise. ((b)) Common information. ((c)) Correlated sensor noise.

Figure 6: The figure illustrates the three main causes of cross-correlations. (a) The process models of the two nodes are affected by the same process noise. (b) An exchange of information means that information is shared by the two nodes. (c) Sensor noise may be correlated if sensors interfere each other.

become

$$\hat{\mathbf{x}}^1_{k+1|k} = \mathbf{F}_k \hat{\mathbf{x}}^1_{k|k} \ , \qquad \hat{\mathbf{x}}^2_{k+1|k} = \mathbf{F}_k \hat{\mathbf{x}}^2_{k|k} \ ,$$

and hence

$$\begin{aligned} \mathbf{P}^{12}_{k+1|k} &= \mathbb{E}[(\hat{\mathbf{x}}^1_{k+1|k} - \mathbf{x}_{k+1})(\hat{\mathbf{x}}^2_{k+1|k} - \mathbf{x}_{k+1})^\mathsf{T}] \\ &= \mathbb{E}\left[(\mathbf{F}_k(\hat{\mathbf{x}}^1_{k|k} - \mathbf{x}_k) + \mathbf{w}_k)(\mathbf{F}_k(\hat{\mathbf{x}}^2_{k|k} - \mathbf{x}_k) + \mathbf{w}_k)^\mathsf{T}\right] \\ &= \mathbf{F}_k \, \mathbb{E}[(\hat{\mathbf{x}}^1_{k|k} - \mathbf{x}_k)(\hat{\mathbf{x}}^2_{k|k} - \mathbf{x}_k)^\mathsf{T}]\mathbf{F}_k^\mathsf{T} + \mathbb{E}[\mathbf{w}_k \mathbf{w}_k^\mathsf{T}] \\ &= \mathbf{F}_k \mathbf{P}^{12}_{k|k} \mathbf{F}_k^\mathsf{T} + \mathbf{Q}_k \\ &= \mathbf{Q}_k \ , \end{aligned}$$

where in the third step it was used that $\mathbb{E}[(\hat{\mathbf{x}}^i_{k|k} - \mathbf{x}_k)\mathbf{w}_k^\mathsf{T}] = \mathbf{0}$. The conclusion is that despite being uncorrelated at time k, the two estimates have become correlated at $k+1$ since the same process noise affects both estimates.

Common Information: One benefit of network-centric data fusion is that information is allowed to be shared, thus improving the accuracy of estimates. Sharing information, however, can lead to correlations since multiple estimates contain the same information after information has been shared. This is the common information problem. For example, assume that we have two estimates $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ with $\mathbf{P}^1 = \mathbb{E}[\tilde{\mathbf{x}}^1(\tilde{\mathbf{x}}^1)^\mathsf{T}]$, $\mathbf{P}^2 = \mathbb{E}[\tilde{\mathbf{x}}^2(\tilde{\mathbf{x}}^2)^\mathsf{T}]$, and $\mathbf{P}^{12} = \mathbb{E}\left[\tilde{\mathbf{x}}^1(\tilde{\mathbf{x}}^2)^\mathsf{T}\right] = \mathbf{0}$. Assume that node 2 sends its estimate to node 1. Since the estimates are uncorrelated, an optimally fused estimate in node 1 is, in this case, given by

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{P}\left((\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + (\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2\right) , \\ \mathbf{P} &= \left((\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1}\right)^{-1} . \end{aligned}$$

After fusion, the correlations become

$$\begin{aligned} \mathbb{E}\left[\tilde{\mathbf{x}}(\tilde{\mathbf{x}}^2)^\mathsf{T}\right] &= \mathbb{E}\left[\mathbf{P}\left((\mathbf{P}^1)^{-1}\tilde{\mathbf{x}}^1 + (\mathbf{P}^2)^{-1}\tilde{\mathbf{x}}^2\right)(\tilde{\mathbf{x}}^2)^\mathsf{T}\right] \\ &= \mathbf{P}(\mathbf{P}^2)^{-1}\,\mathbb{E}\left[\tilde{\mathbf{x}}^2(\tilde{\mathbf{x}}^2)^\mathsf{T}\right] \\ &= \mathbf{P}(\mathbf{P}^2)^{-1}\mathbf{P}^2 \\ &= \mathbf{P} \ . \end{aligned}$$

The fused estimate at node 1 is now correlated with the estimate at node 2. The actual information that is common to $\hat{\mathbf{x}}$ and

$\hat{\mathbf{x}}^2$, in this case, is $(\mathbf{P}^2)^{-1}$ due to $\mathbf{P}^{-1} = (\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1}$. Common information is not exclusively due to fusion of estimates. Similar expressions, for example, are obtained if both nodes access the same sensor, and $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ would be updated with the same measurement.

Double counting [34] of information is intuitively visualized using common information. Assume that $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ are

$$(\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 = \mathbf{A}^{-1}\hat{\mathbf{a}} + \mathbf{\Gamma}^{-1}\hat{\boldsymbol{\gamma}} \ , \quad (\mathbf{P}^1)^{-1} = \mathbf{A}^{-1} + \mathbf{\Gamma}^{-1} \tag{15a}$$

$$(\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2 = \mathbf{B}^{-1}\hat{\mathbf{b}} + \mathbf{\Gamma}^{-1}\hat{\boldsymbol{\gamma}} \ , \quad (\mathbf{P}^2)^{-1} = \mathbf{B}^{-1} + \mathbf{\Gamma}^{-1} \tag{15b}$$

where $\hat{\boldsymbol{\gamma}}$ with $\mathbf{\Gamma}$ is the common information shared by both estimates, while $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\boldsymbol{\gamma}}$ are mutually uncorrelated. Then, naïve fusion of $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ results in

$$\mathbf{P}^{-1}\hat{\mathbf{x}} = (\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + (\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2 = \mathbf{A}^{-1}\hat{\mathbf{a}} + \mathbf{B}^{-1}\hat{\mathbf{b}} + 2\mathbf{\Gamma}^{-1}\hat{\boldsymbol{\gamma}} \ ,$$

$$\mathbf{P}^{-1} = (\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1} + 2\mathbf{\Gamma} \ ,$$

where $\hat{\boldsymbol{\gamma}}$ with $\mathbf{\Gamma}$ is counted twice. The convex combination (8) directly shows how CI avoids double counting [35]. Hence, CI incorporates $\hat{\boldsymbol{\gamma}}$ with $\mathbf{\Gamma}$ only once.

An important aspect of devising fusion methods is that common process noise and common information cannot be treated as separate entities if they are simultaneously present. For instance, common information $\hat{\boldsymbol{\gamma}}_{k|k}$ in (15) at a time step k gets blended with the process noise $\mathbf{w}_k$ after prediction, which leads to

$$\mathbf{P}^1_{k+1|k} = \mathbf{F}_k\mathbf{P}^1_{k|k}\mathbf{F}_k^\mathsf{T} + \mathbf{Q}_k = \mathbf{F}_k\left(\mathbf{A}^{-1}_{k|k} + \mathbf{\Gamma}^{-1}_{k|k}\right)^{-1}\mathbf{F}_k^\mathsf{T} + \mathbf{Q}_k$$

and corresponding $\mathbf{P}^2_{k+1|k}$. Hence, common information from former time steps cannot be represented in the form (15) anymore. An implication is that fusion methods cannot treat process noise and common information separately after multiple time steps.

Correlated Sensor Noise: A common assumption in data fusion is that the noises from different sensors are independent. In some scenarios, this assumption cannot be guaranteed [36], [37], and cross-correlated sensor noise has to be taken into consideration. Consider two uncorrelated estimates $\hat{\mathbf{x}}^1_{k|k-1}$ and $\hat{\mathbf{x}}^2_{k|k-1}$, that is, $\mathbf{P}^{12}_{k|k-1} = \mathbb{E}[\tilde{\mathbf{x}}^1_{k|k-1}(\tilde{\mathbf{x}}^2_{k|k-1})^\mathsf{T}] = \mathbf{0}$. For each estimate, a Kalman filter update is performed with a measurement $\mathbf{y}^i_k = \mathbf{H}^i_k\mathbf{x}_k + \mathbf{v}^i_k$, $i = 1, 2$, where $\mathbf{H}^i$ is the measurement mapping and $\mathbf{v}^i$ is the zero-mean sensor noise with $\mathbf{R}^i_k = \mathbb{E}[\mathbf{v}^i_k(\mathbf{v}^i_k)^\mathsf{T}]$. A typical assumption is that the sensor noises are uncorrelated. However, if both sensors interfere with each other, nonzero correlations $\mathbf{R}^{12}_k = \mathbb{E}[\mathbf{v}^1_k(\mathbf{v}^2_k)^\mathsf{T}] \neq \mathbf{0}$ will be present. The corresponding Kalman updates yield

$$(\mathbf{P}^i_{k|k})^{-1}\hat{\mathbf{x}}^i_{k|k} = (\mathbf{P}^i_{k|k-1})^{-1}\hat{\mathbf{x}}^i_{k|k-1} + (\mathbf{H}^i_k)^\mathsf{T}(\mathbf{R}^i_k)^{-1}\mathbf{y}^i_k \ ,$$

$$(\mathbf{P}^i_{k|k})^{-1} = (\mathbf{P}^i_{k|k-1})^{-1} + (\mathbf{H}^i_k)^\mathsf{T}(\mathbf{R}^i_k)^{-1}\mathbf{H}^i_k$$

and entail the cross-covariance structure

$$\begin{aligned}
&\mathbb{E}\left[\tilde{\mathbf{x}}^1_{k|k}(\tilde{\mathbf{x}}^2_{k|k})^\mathsf{T}\right] \\
&\quad = \mathbb{E}\Big[\mathbf{P}^1_{k|k}\left((\mathbf{P}^1_{k|k-1})^{-1}\tilde{\mathbf{x}}^1_{k|k-1} + (\mathbf{H}^1_k)^\mathsf{T}(\mathbf{R}^1_k)^{-1}\mathbf{v}^1_k\right) \\
&\qquad \times \left((\mathbf{P}^2_{k|k-1})^{-1}\tilde{\mathbf{x}}^2_{k|k-1} + (\mathbf{H}^2_k)^\mathsf{T}(\mathbf{R}^2_k)^{-1}\mathbf{v}^2_k\right)^\mathsf{T}\mathbf{P}^2_{k|k}\Big] \\
&= \mathbb{E}\left[\mathbf{P}^1_{k|k}(\mathbf{H}^1_k)^\mathsf{T}(\mathbf{R}^1_k)^{-1}\mathbf{v}^1_k(\mathbf{v}^2_k)^\mathsf{T}(\mathbf{R}^2_k)^{-1}\mathbf{H}^2_k\mathbf{P}^2_{k|k}\right] \\
&= \mathbf{P}^1_{k|k}(\mathbf{H}^1_k)^\mathsf{T}(\mathbf{R}^1_k)^{-1}\mathbf{R}^{12}_k(\mathbf{R}^2_k)^{-1}\mathbf{H}^2_k\mathbf{P}^2_{k|k} \ .
\end{aligned}$$

Hence, correlated sensor noise causes correlations among $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ even though they are processed on independent nodes. This third case can be viewed as a generalization of common information if $\mathbf{H}^i = \mathbf{I}$, $\mathbf{R}^i = \mathbf{\Gamma}$, and $\mathbf{v}^i = \hat{\boldsymbol{\gamma}}$ for $i = 1, 2$.

Consequences of Cross-Correlations

For a practical example of what happens when cross-correlations are neglected, consider the simulated scenario depicted in Figure 7. Three vehicles approach a three-way junction. The target vehicle of interest (white van) follows the dashed trajectory intends to turn left at the junction. The orange truck and blue car, each equipped with a sensor and an EKF, track the target vehicle. To further improve their estimates, they exchange their local track estimates and fuse them.

The estimates are fused using naïve fusion where cross-correlations are neglected and using CI based on the same noise realization. The results are shown in Figure 8, where only estimates computed in the blue car are shown. Initially the naïve fusion rule performs relatively well but as time progresses and double counting of information starts to dominate, the naïve fusion estimate (yellow) becomes too optimistic. The consequence is that the estimate overshoots the turn of the target. Meanwhile, the same effect is not seen in case of CI (green) which is robust to double counting of information. An estimate which only uses local information is provided for comparison (blue).

The specific results of Figure 8 is only one simulated example. Nevertheless, these results are typical and characterize the effect of neglecting cross-correlations by double counting information. That is, the naïve estimator becomes too confident of the estimate and hence becomes partly incapable of incorporating new measurements.

Optimality of Conservative Fusion

An important question is whether and in which regard CI is an optimal fusion rule. Optimality typically refers to known cross-covariances $\mathbf{P}^{12}$, which enable the computation of the fusion result (4). In the special case $\mathbf{P}^{12} = \mathbf{0}$, optimal fusion simply yields the fused covariance matrix

$$\mathbf{P}^{-1} = (\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1} \ ,$$

which equals naïve fusion. If we instead apply CI, we obtain far too conservative results $\mathbf{P}^{\mathrm{CI}} \succ \mathbf{P}$ as the gap between the green and red ellipse in Figure 4 indicates. This is no surprise, CI is derived for the case when $\mathbf{P}^{12}$ is unknown. In the case of completely unknown $\mathbf{P}^{12}$, it however turns out that CI is indeed optimal.

To study the optimality of CI, we need

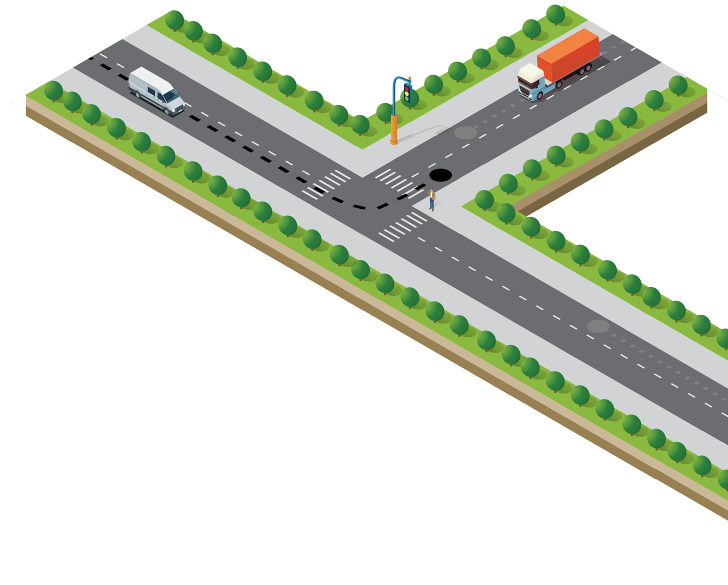


Figure 7: Scenario used to illustrate the effect of neglecting cross-correlations. A target vehicle, the white van, is tracked by sensors located in two other vehicles, the blue car and the orange truck. All vehicles follow the dashed trajectories and approach the three-way junction.

to scrutinize the corresponding error

$$\tilde{\mathbf{x}}^{\mathrm{CI}} = \hat{\mathbf{x}}^{\mathrm{CI}} - \mathbf{x} = \mathbf{K}^{\mathrm{CI}}\,\tilde{\mathbf{x}}^1 + \mathbf{L}^{\mathrm{CI}}\,\tilde{\mathbf{x}}^2$$

for the CI estimate in (5a) with

$$\mathbf{K}^{\mathrm{CI}} = \omega \mathbf{P}^{\mathrm{CI}}(\mathbf{P}^1)^{-1}\ , \quad \mathbf{L}^{\mathrm{CI}} = (1-\omega)\mathbf{P}^{\mathrm{CI}}(\mathbf{P}^2)^{-1}\ ,$$

and $\mathbf{I} = \mathbf{K}^{\mathrm{CI}} + \mathbf{L}^{\mathrm{CI}}$. The actual error covariance matrix then becomes

$$\begin{aligned}\mathbb{E}\left[\tilde{\mathbf{x}}^{\mathrm{CI}}(\tilde{\mathbf{x}}^{\mathrm{CI}})^\top\right] &= \mathbf{K}^{\mathrm{CI}}\mathbf{P}^1(\mathbf{K}^{\mathrm{CI}})^\top + \mathbf{K}^{\mathrm{CI}}\mathbf{P}^{12}(\mathbf{L}^{\mathrm{CI}})^\top \\ &\quad + \mathbf{L}^{\mathrm{CI}}(\mathbf{P}^{12})^\top(\mathbf{K}^{\mathrm{CI}})^\top + \mathbf{L}^{\mathrm{CI}}\mathbf{P}^2(\mathbf{L}^{\mathrm{CI}})^\top ,\end{aligned} \tag{16}$$

which is unknown to the fusion algorithm because $\mathbf{P}^{12}$ is unknown. However, Figure 9 reveals that CI tightly circumscribes the union of all ellipses $\mathcal{E}\left(\mathbb{E}\left[\tilde{\mathbf{x}}^{\mathrm{CI}}(\tilde{\mathbf{x}}^{\mathrm{CI}})^\top\right]\right)$. This illustration provides an intuition for CI's optimality. Note that CI uses different gains $\mathbf{K}^{\mathrm{CI}}$ and $\mathbf{L}^{\mathrm{CI}}$ than BC. Therefore, Figure 4 showing the ellipses for BC neither illustrates conservativeness nor optimality of CI while Figure 9 shows $\mathbf{P}^{\mathrm{CI}}$ as bound on the possible actual errors (16).

As noted in [21], [29] $\mathbf{P}^{\mathrm{CI}}$ is the smallest covariance for any estimate computed linearly from $\hat{\mathbf{x}}^J$ with covariance $\mathbf{P}^{J,\mathrm{CI}}$ according to (13). However, as pointed out in [38], this is not sufficient for proving CI is optimal. A key result is given in [38] showing that there is no tighter bound on $\mathbb{E}[\tilde{\mathbf{x}}(\tilde{\mathbf{x}})^\top]$ than $\mathbf{P}^{\mathrm{CI}}$ if $\mathbf{P}^{12}$ is completely unknown. More specifically, they show that CI is optimal

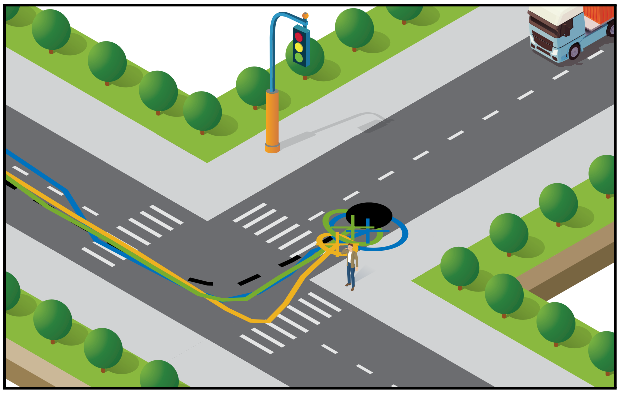


Figure 8: Results of the neglecting cross-correlations scenario. The estimates are fused using naïve fusion and covariance intersection. The results of only using local information are also plotted. The estimated trajectories of the naïve fusion, covariance intersection and local information only are given by the yellow, the green and the blue lines, respectively. The double counting of information present in case of naïve fusion produces an estimate having a too optimistic covariance with the consequence that the estimate overshoots the turn of the target.

with respect to a strictly monotonically increasing loss function J. The work in [39] suggests focusing on the trace of the fused covariance matrix. In doing so, a smaller covariance matrix can be computed by bounding $\operatorname{tr}\left(\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\intercal]\right)$ instead of $\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\intercal]$ and, hence, the trace of $\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\intercal]$ is bounded directly. However, with this tighter bound, it follows that the conservativeness criterion may be violated.

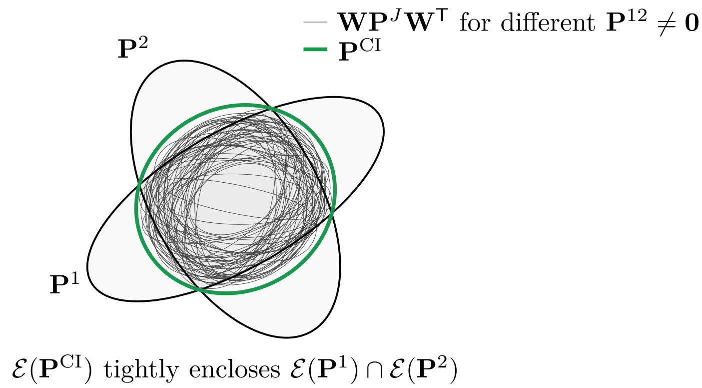


Figure 9: Geometrical interpretation of optimality of covariance intersection. Let $\mathbf{W} = \begin{bmatrix}\mathbf{K}^{\mathrm{CI}} & \mathbf{L}^{\mathrm{CI}}\end{bmatrix}$. It is known from the subsequent section that the ellipses $\mathcal{E}(\mathbf{P}^{\mathrm{BC}})$ fills the intersection $\mathcal{E}(\mathbf{P}^1) \cap \mathcal{E}(\mathbf{P}^2)$. Hence, since CI tightly encloses $\mathcal{E}(\mathbf{P}^1) \cap \mathcal{E}(\mathbf{P}^2)$ and at the same time $\mathcal{E}(\mathbf{P}^{\mathrm{CI}}) \supseteq \mathcal{E}(\mathbf{W}\mathbf{P}^J\mathbf{W}^\intercal)$ for all admissible $\mathbf{P}^J$, CI is an optimal conservative estimator.

Computational Aspects and Multiple Estimates

While the optimal fusion in (4) is given in closed form when correlations are known, the fusion result of CI in (5) depends on the parameter ω. The fusion result of CI forms a family of possible solutions (5), and the optimal choice of ω typically requires a numerical approach when the trace or determinant is the considered cost function. This computational overhead has triggered further research on new approaches to determining the weight parameter ω, and particular attention is required when multiple estimates are fused using CI, either sequential or batchwise.

Computation of the Weight Parameters

We have seen that CI provides a family of estimates as a consequence of the free parameter $\omega \in [0,1]$. While each $\omega \in [0,1]$ ensures a conservative estimate, for the performance, it is however crucial to compute an ω that yields a $\mathbf{P}^{\mathrm{CI}}$ which is as small as possible with respect to $J(\mathbf{P}^{\mathrm{CI}})$. This task might be difficult, computationally demanding, or even intractable, given the actual system design. Below we survey several techniques that address this.

In [40], different principles to determine ω are studied, namely, Shannon fusion and Bhattacharyya fusion. Shannon fusion refers to minimizing $\det(\mathbf{P}^{\mathrm{CI}})$ and is, in the Gaussian case, the same as maximization of the Shannon information. This criterion is justified by information theory [41]. In Bhattacharyya fusion, ω is simply chosen as $\omega = 1/2$, and (5) hence reduces to

$$(\mathbf{P}^{\mathrm{CI}})^{-1}\hat{\mathbf{x}}^{\mathrm{CI}} = \frac{1}{2}\left((\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + (\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2\right) \ ,$$
$$(\mathbf{P}^{\mathrm{CI}})^{-1} = \frac{1}{2}\left((\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1}\right) \ ,$$

which resemble naïve fusion (6) except for the factor 1/2. This means that Bhattacharyya fusion yields the same estimate $\hat{\mathbf{x}}^{\mathrm{CI}}$ as naïve fusion but with a twice as large covariance matrix $\mathbf{P}^{\mathrm{CI}}$. Such a simple choice does not take into account the different contributions of the two estimates, and therefore might lead to undesirable results. For instance, consider

$$\mathbf{P}^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{P}^2 = \begin{bmatrix} 10 & 0 \\ 0 & 0.5 \end{bmatrix},$$

and $\mathrm{tr}(\cdot)$ as the loss function. Then, with the trace optimal ω we get $\mathrm{tr}(\mathbf{P}^{\mathrm{CI}})/\,\mathrm{tr}(\mathbf{P}^1) < 1$ which is a strict improvement. However, with $\omega = 1/2$ we get $\mathrm{tr}(\mathbf{P}^{\mathrm{CI}})/\,\mathrm{tr}(\mathbf{P}^1) > 1$ which actually is a performance degradation, with respect to trace, compared to $\mathbf{P}^1$.

A *fast CI* algorithm is proposed in [42], which provides closed-form expressions for an approximate computation of ω based on the individual covariance matrices. Fast CI is based on a linear constraint weighted by the trace of covariance matrices

$$\mathrm{tr}(\mathbf{P}^1)\omega - \mathrm{tr}(\mathbf{P}^2)(1-\omega) = 0 \ ,$$

such that

$$\omega = \frac{\mathrm{tr}(\mathbf{P}^2)}{\mathrm{tr}(\mathbf{P}^1) + \mathrm{tr}(\mathbf{P}^2)} \ . \tag{17}$$

Hence, fast CI utilizes a trace-based fusion gain similar to the scalar gains computed in [43].

An improved version of fast CI based on determinant weighting is proposed in [44].

The fast CI of [45], [46] relies upon Kullback-Leibler divergence and is in fact robust even in the case when one of the estimates to be merged is not conservative. The latter is accomplished by computing the weight parameter ω in a subinterval $[0, \delta] \subseteq [0, 1]$, where $0 < \delta \leq 1$ is derived adaptively to preserve conservativeness, and then replace the weight $(1 - \omega)$ corresponding to $(\hat{\mathbf{x}}^2, \mathbf{P}^2)$ by $(\delta - \omega)$. The closed-form expressions derived in the fast CI algorithm enable CI to be deployed in real-world systems to cope with implementations challenges: In [47] it is shown that the fast CI algorithm of [42] can be computed homomorphically for encrypted estimates. The resulting method is a secure fast CI fusion scheme which is further pursued in [48]. The problem of conservative fusion of quantized estimates is studied in [49], [50]. A benefit of data compression by quantization is to reduce the communication load which is a current topic in DDF. For other examples of where CI is used in a DDF under communication constraints, see [51] where the exchanged covariances are approximated by diagonal matrices, and [52]–[54] where dimension-reduced estimates are exchanged.

If J is $\mathrm{tr}(\cdot)$ or $\det(\cdot)$, then the problem in (9) is a convex optimization problem [22], which implies desirable convergence properties when it comes to optimization of ω. For unimodal J, for example, if J is strongly convex, simple optimization algorithms such as the golden section search [55] can be used. In [56], closed-form expressions are derived for low state dimensionalities n. The results

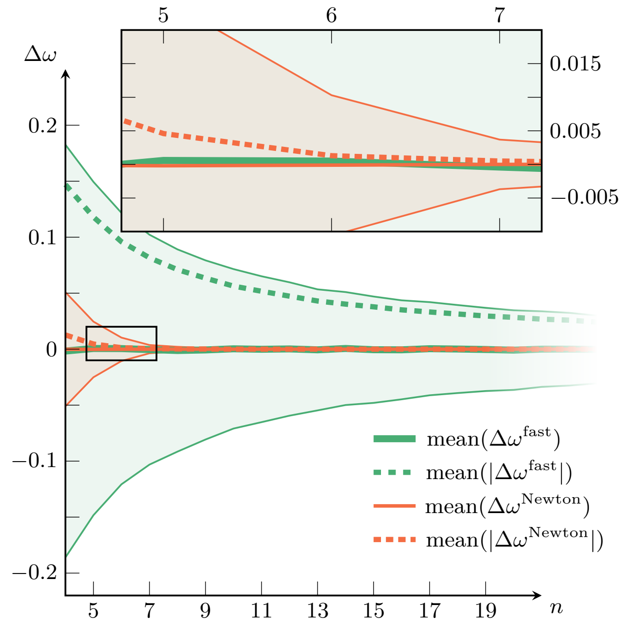


Figure 10: Comparison of weight parameter computation methods. Fast CI and Newton CI approximations for computation of ω are based on sampling of $\mathbf{P}^1$ and $\mathbf{P}^2$ and trace optimization. The function $\mathrm{mean}(\cdot)$ denotes sample mean and n is the dimensionality. The performance of Newton CI is superior the fast CI algorithm.

are exact, and the formulas are valid for $n \leq 3$ and $n \leq 4$ in case of trace and determinant, respectively. For larger dimensions, the results of [22] can be used, in which two iterations of Newton's method yield approximative closed-form expressions. The approximative expressions are applicable for both trace and determinant minimization. This *Newton CI* method effectively approximates the optimal ω for entropy and variance cost functions. It reaches close-to-optimal results with comparably low computational costs, which is demonstrated in the remainder of this section.

The weight parameter ω is derived based on trace optimization, and $\mathbf{P}^1$ and $\mathbf{P}^2$ are covariance matrices sampled from a Wishart distribution $W(I, n)$ [57]. Let ω^{fast} be computed according to (17) and ω^{Newton} be computed according to the trace optimization approximation in [22]. Moreover, let $\Delta\omega^{\text{Newton}} = \omega^{\text{fast}} - \omega$ and $\Delta\omega^{\text{Newton}} = \omega^{\text{Newton}} - \omega$, where ω is the true value computed as the solution to (9). For different values of n, $\Delta\omega^{\text{fast}}$ and $\Delta\omega^{\text{Newton}}$ are computed for randomly generated $\mathbf{P}^1$ and $\mathbf{P}^2$. For each n a number of 100 000 samples are generated. The sample means of $\Delta\omega^{\text{fast}}$ and $\Delta\omega^{\text{Newton}}$, and $|\Delta\omega^{\text{fast}}|$ and $|\Delta\omega^{\text{Newton}}|$, are illustrated in Figure 10. The shaded areas visualized the sample standard deviations of $\Delta\omega^{\text{fast}}$ and $\Delta\omega^{\text{Newton}}$. The Newton CI algorithm clearly outperforms fast CI. Both approximations become more exact as n increases. It is also seen that both $\Delta\omega^{\text{fast}}$ and $\Delta\omega^{\text{Newton}}$ are zero-mean.

Fusion of Multiple Estimates

To merge $N \geq 2$ estimates $\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \ldots, \hat{\mathbf{x}}^N$ it is always possible to subsequently apply the CI formulas of (8). For example, consider fusion of $\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2$ and $\hat{\mathbf{x}}^3$ where

$$\mathbf{P}^1 = \begin{bmatrix} 12 & 0 \\ 0 & 3 \end{bmatrix}, \quad \mathbf{P}^2 = \begin{bmatrix} 12 & -5 \\ -5 & 5 \end{bmatrix}, \quad \mathbf{P}^3 = \begin{bmatrix} 4 & 3 \\ 3 & 10 \end{bmatrix}. \tag{18}$$

We can now, in a first step, fuse $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$, and then, in a second step, fuse the results of the former step with $\hat{\mathbf{x}}^3$, both using (8). Let

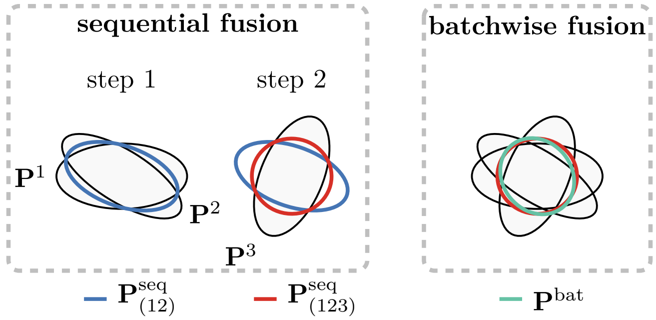


Figure 11: Fusion of multiple estimates using covariance intersection. Sequential fusion is compared to batchwise fusion of three estimates. In the sequential update $\mathbf{P}^1$ and $\mathbf{P}^2$ are fused into an intermediate covariance matrix $\mathbf{P}^{\text{seq}}_{(12)}$ which is then fused with $\mathbf{P}^3$. The batch method fused $\mathbf{P}^1$, $\mathbf{P}^2$ and $\mathbf{P}^3$ simultaneously. In this example batchwise fusion yields a smaller covariance with respect to $\text{tr}(\cdot)$ compared sequential fusion.

$$\mathbf{P}^{\text{seq}}_{(12)} = \left(\omega^{\text{seq}}_{(12)}(\mathbf{P}^1)^{-1} + (1 - \omega^{\text{seq}}_{(12)})(\mathbf{P}^2)^{-1}\right)^{-1} = \begin{bmatrix} 8.80 & -2.20 \\ -2.20 & 3.30 \end{bmatrix}$$

and

$$\mathbf{P}^{\text{seq}}_{(123)} = \left(\omega^{\text{seq}}_{(123)}(\mathbf{P}^{\text{seq}}_{(123)})^{-1} + (1 - \omega^{\text{seq}}_{(123)})(\mathbf{P}^3)^{-1}\right)^{-1} = \begin{bmatrix} 4.43 & -0.02 \\ -0.02 & 3.99 \end{bmatrix},$$

where $\omega^{\text{seq}}_{(12)}$ and $\omega^{\text{seq}}_{(123)}$ were computed based on trace minimization in each step. The results are illustrated in Figure 11. Since this sequential approach involves optimizing for ω in two steps and neither of the steps takes into account all three estimates, the final result is in general suboptimal compared to optimizing ω for all three estimates simultaneously. The best way is therefore to fuse all three estimates batchwise.

Batchwise fusion using CI is formulated as follows for the general case. Assume that N estimates of $\mathbf{x}$ are available and that the ith estimate is $\hat{\mathbf{x}}^i = \mathbf{x} + \tilde{\mathbf{x}}^i$. The estimation error $\tilde{\mathbf{x}}^i$ has the covariance matrix $\mathbf{P}^i = \mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}]$. CI for batchwise fusion of $N \geq 2$ estimates has the form

$$(\mathbf{P}^{\mathrm{CI}})^{-1}\hat{\mathbf{x}}^{\mathrm{CI}} = \sum_{i=1}^{N} \omega_i (\mathbf{P}^i)^{-1}\hat{\mathbf{x}}^i \ , \quad (19a)$$

$$(\mathbf{P}^{\mathrm{CI}})^{-1} = \sum_{i=1}^{N} \omega_i (\mathbf{P}^i)^{-1} \ , \quad (19b)$$

where $\omega_i \in [0,1]$ and $\sum_{i=1}^{N} \omega_i = 1$. Fusion of multiple estimates can also be expressed in the joint form, which was introduced for $N = 2$ in (10). In this case, the joint estimate and joint covariance are given by

$$\hat{\mathbf{x}}^J = \begin{bmatrix} \hat{\mathbf{x}}^1 \\ \hat{\mathbf{x}}^2 \\ \vdots \\ \hat{\mathbf{x}}^N \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \tilde{\mathbf{x}}^1 \\ \tilde{\mathbf{x}}^2 \\ \vdots \\ \tilde{\mathbf{x}}^N \end{bmatrix} = \mathbf{H}^J\mathbf{x} + \tilde{\mathbf{x}}^J \ ,$$

$$\mathbf{P}^J = \begin{bmatrix} \mathbf{P}^1 & \mathbf{P}^{12} & \dots & \mathbf{P}^{1N} \\ \mathbf{P}^{21} & \mathbf{P}^2 & \dots & \mathbf{P}^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}^{N1} & \mathbf{P}^{N2} & \dots & \mathbf{P}^N \end{bmatrix} = \mathbb{E}[(\tilde{\mathbf{x}}^J)(\tilde{\mathbf{x}}^J)^\mathsf{T}] \ .$$

A conservative bound for actual but unknown $\mathbf{P}^J$ is

$$\mathbf{P}^{J,\mathrm{CI}} = \begin{bmatrix} \frac{\mathbf{P}^1}{\omega_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{P}^2}{\omega_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \frac{\mathbf{P}^N}{\omega_N} \end{bmatrix} \succeq \mathbf{P}^J \ , \quad (20)$$

corresponding to (12). Analogously to (13), the CI formulas in (19) can then be expressed as

$$(\mathbf{P}^{\mathrm{CI}})^{-1}\hat{\mathbf{x}}^{\mathrm{CI}} = (\mathbf{H}^J)^\mathsf{T}(\mathbf{P}^{J,\mathrm{CI}})^{-1}\hat{\mathbf{x}}^J \ , \quad (21a)$$

$$(\mathbf{P}^{\mathrm{CI}})^{-1} = (\mathbf{H}^J)^\mathsf{T}(\mathbf{P}^{J,\mathrm{CI}})^{-1}\mathbf{H}^J \ , \quad (21b)$$

which are a weighted least squares (WLS) representation of CI for multiple estimates.

In the previous example with $\mathbf{P}^1$, $\mathbf{P}^2$ and $\mathbf{P}^3$ having the parameter (18), we obtain

$$\mathbf{P}^{\mathrm{bat}} = \left(\omega_1(\mathbf{P}^1)^{-1} + \omega_2(\mathbf{P}^2)^{-1} + \omega_3(\mathbf{P}^3)^{-1}\right)^{-1}$$

$$= \begin{bmatrix} 3.83 & -0.58 \\ -0.58 & 4.05 \end{bmatrix} ,$$

by using batchwise fusion (19), where ω_1, ω_2 and ω_3 are found by trace minimization. Since $\mathrm{tr}(\mathbf{P}^{\mathrm{seq}}) \approx 8.41$ and $\mathrm{tr}(\mathbf{P}^{\mathrm{bat}}) \approx 7.89$ we have

$$\mathrm{tr}(\mathbf{P}^{\mathrm{seq}}) > \mathrm{tr}(\mathbf{P}^{\mathrm{bat}}) \ ,$$

and hence batchwise fusion is superior to sequential fusion. The results are illustrated in the right part of Figure 11.

In [28], it is shown that CI produces conservative estimates also if $N > 2$. Fusion of multiple estimates simultaneously does, however, not preserve optimality. The main downside of using CI batchwise is the increased computational complexity due to replacing the 1D problem of optimizing for ω with the multidimensional problem of determining $\omega_1, \omega_2, \dots, \omega_{N-1}$, where ω_N is excluded since it is given by $\omega_N = 1 - \sum_{i=1}^{N-1} \omega_i$. A comprehensive analysis of batch CI compared to sequentially applying CI to two estimates is provided in [58].

Finding the optimal weighting parameters is even more challenging for the fusion of multiple estimates. One advantage of fast CI is that it is order-invariant and also applicable for batchwise fusion. In the $N \geq 2$ case, there are $N-1$ constraints

$$\mathrm{tr}(\mathbf{P}^i)\omega_i - \mathrm{tr}(\mathbf{P}^{i+1})\omega_{i+1} = 0 \,, \quad i = 1, 2, \ldots, N-1 \,,$$

which, using $\varsigma_i = \mathrm{tr}(\mathbf{P}^i)$ and including $\sum_{i=1}^{N} \omega_i = 1$, yield the linear system of equations

$$\begin{bmatrix} \varsigma_1 & -\varsigma_2 & 0 & \ldots & 0 \\ 0 & \varsigma_2 & -\varsigma_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \varsigma_{N-1} & -\varsigma_N \\ 1 & \ldots & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_{N-1} \\ \omega_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} . \tag{22}$$

Solving (22) finally yields

$$\omega_i = \frac{\frac{1}{\varsigma_i}}{\sum_{j=1}^{N} \frac{1}{\varsigma_j}} = \frac{\frac{1}{\mathrm{tr}(\mathbf{P}^i)}}{\sum_{j=1}^{N} \frac{1}{\mathrm{tr}(\mathbf{P}^j)}} \,,$$

which shows that fast solutions for the fusion of multiple estimates are also attainable.

Other Practical Aspects

In some cases, the choice of a loss function $J(\mathbf{P}^{\mathrm{CI}})$, like the trace or determinant, may lead to an undesirable jumping between estimates. As discussed earlier in section *Problem Intuition*, CI chooses the estimate with a smaller covariance matrix when $\mathbf{P}^1 \succeq \mathbf{P}^2$ or $\mathbf{P}^2 \succeq \mathbf{P}^1$ holds. If this relationship switches, for example, every other time step, the fusing result will jump accordingly between the two estimates. This is, in particular, a problem in one dimension and is an argument for considering different criteria for choosing ω in such cases. One alternative is to consider set-membership methods. A set-theoretic interpretation of CI was originally given in [59]. By associating information with ellipsoids, [46], [60] propose to employ the Chebyshev center to compute optimal weights $\omega_1, \ldots, \omega_N$ for $N \geq 2$ estimates. In [61], the set of possible correlation coefficients is considered and studied. In a subsequent step, the same coefficient is assumed to be uniformly distributed and is marginalized out to obtain a fusion result.

Especially in object tracking scenarios, CI needs to handle fusion of estimates over multiple time steps in a dynamic environment. This is not an issue since any cross-correlations originating from common process noise, sharing of information, or any other source are handled conservatively by CI. However, recursive application of CI over time may lead to unnecessarily large covariances as shown in [62]. Therefore, MSE optimality is not preserved when using CI recursively in dynamic systems [63]. Several fusion schemes involving CI have been developed to better preserve information over time. The algorithms in [64] and [65] are based upon fusion without and with memory, respectively, and feedback mechanisms to more efficiently use information in dynamic systems. Different robust fusion mechanisms for merging of multiple estimates in time-varying systems are studied in [66], [67], where CI is compared to different fusion techniques using the joint covariance.

From a user's perspective, *Sidebar: Practitioner's Point-Of-View* provides a guideline and discusses how to find a proper choice of a fusion method. Alternatives and derivatives of CI are the subject of the following section.

Derivatives and Alternatives to Covariance Intersection

CI is conservative for all possible cross-correlations. When two estimates have completely unknown cross-correlations, CI yields the optimal fusion result in the sense that it has the tightest bound on the actual error. However, partial knowledge about the cross-correlations or their sources is available in many situations. In such cases, CI becomes overly conservative, and less conservative alternatives should be considered instead. The recent study [68] further discusses fusion under partially known cross-correlations and introduces the notion of conservative linear unbiased estimation (CLUE). In essence, the methods described in the following are instances of a CLUE. In this context, CI is an optimal CLUE if two estimates are considered and $\mathbf{P}^{12}$ is completely unknown [38]. The following methods are optimal under different assumptions on partially known cross-correlations. Figure 12 provides a coarse classification of the underlying assumptions that allow for less conservative fusion results. In the following, we highlight those significant contributions in a chronological order.

Federated Kalman Filter

The federated Kalman filter (FKF, [69], [70]) was proposed a decade before CI. It can however be regarded as a special case of CI tailored to cope with common process noise. The FKF accomplishes this by inflating the joint covariance. It assumes that N local filters are initialized with the same estimate $\hat{\mathbf{x}}_{0|0}$, which has the covariance matrix $\mathbf{P}_{0|0}$. Hence, the local estimates are fully correlated, that is, $\mathbb{E}[\tilde{\mathbf{x}}^i_{0|0}(\tilde{\mathbf{x}}^j_{0|0})^\top] = \mathbf{P}_{0|0}$. The FKF is a centralized scheme, in which each node has a local filter processing local estimates, and a central node collects and fuses the local results. The central node can also reinitialize the local filters with the fusion result. Local measurements are assumed to be uncorrelated, but each filter is affected by the same process noise in the prediction step as discussed in *Disturbing Disturbances*. To address correlations, the FKF makes the following adaptations to the local filters: The initialization at $k = 0$ or reinitialization at a time step k become

$$\hat{\mathbf{x}}^i_{k|k} = \hat{\mathbf{x}}_{k|k}, \qquad (\mathbf{P}^i_{k|k})^{-1} = \beta_i \mathbf{P}^{-1}_{k|k} \ , \tag{23}$$

and the prediction step is altered to

$$\begin{aligned} \hat{\mathbf{x}}^i_{k+1|k} &= \mathbf{F}_k \, \hat{\mathbf{x}}^i_{k|k} \ , \\ \mathbf{P}^i_{k+1|k} &= \mathbf{F}_k \, \mathbf{P}^i_{k|k} \, \mathbf{F}_k^\top + \frac{1}{\beta_i} \mathbf{Q}_k \end{aligned} \tag{24}$$

for each node $i = 1, \ldots, N$ with $\beta_i > 0, \sum_{i=1}^n \beta_i = 1$. More precisely, an inflation parameter β_i has been introduced in the covariance matrix equations, which corresponds to the bound

$$\mathbb{E}\left[\tilde{\mathbf{x}}^J \left(\tilde{\mathbf{x}}^J\right)^\top\right] = \begin{bmatrix} \mathbf{P}^i_{k|k} & \cdots & \mathbf{P}^i_{k|k} \\ \vdots & \ddots & \vdots \\ \mathbf{P}^i_{k|k} & \cdots & \mathbf{P}^i_{k|k} \end{bmatrix} \preceq \begin{bmatrix} \frac{1}{\beta_1}\mathbf{P}^i_{k|k} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \frac{1}{\beta_N}\mathbf{P}^i_{k|k} \end{bmatrix}$$

on the joint error covariance matrix for the (re-)initialization (23) of the local filers, where $\tilde{\mathbf{x}}^J = [(\tilde{\mathbf{x}}^1)^\top, \ldots, (\tilde{\mathbf{x}}^N)^\top]^\top$ is the joint estimation error. For the common process noise, the FKF uses the bound

$$\mathbb{E}\left[\mathbf{w}_k^J(\mathbf{w}_k^J)^\top\right] = \begin{bmatrix}\mathbf{Q}_k & \cdots & \mathbf{Q}_k \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_k & \cdots & \mathbf{Q}_k\end{bmatrix} \preceq \begin{bmatrix}\frac{1}{\beta_1}\mathbf{Q}_k & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \frac{1}{\beta_N}\mathbf{Q}_k\end{bmatrix}$$

with $\mathbf{w}_k^J = [\mathbf{w}_k^\top, \ldots, \mathbf{w}_k^\top]^\top$, which justifies the inflated covariance matrices in (24). The inflation is the same technique CI uses for the inflated joint covariance matrix (20). The FKF assumes a central dedicated fusion center, although generalizations are possible. However, it does not consider the problem of common information or correlated sensor noise.

Split CI and Bounded Correlations

Split CI is based on the assumption that each estimate can be split into an independent part and a dependent part [3], [10], [71], [72]. The estimates are assumed to have the form $\hat{\mathbf{x}}^1 = \hat{\mathbf{x}}_e^1 + \hat{\mathbf{x}}_d^1$ and $\hat{\mathbf{x}}^2 = \hat{\mathbf{x}}_e^2 + \hat{\mathbf{x}}_d^2$, where $\hat{\mathbf{x}}_e^1$ and $\hat{\mathbf{x}}_e^2$ are uncorrelated, and $\hat{\mathbf{x}}_d^1$ and $\hat{\mathbf{x}}_d^2$ have unknown correlations. In other words, $\mathbb{E}[\tilde{\mathbf{x}}_e^1(\tilde{\mathbf{x}}_e^2)^\top] = \mathbb{E}[\tilde{\mathbf{x}}_d^1(\tilde{\mathbf{x}}_e^2)^\top] = \mathbb{E}[\tilde{\mathbf{x}}_e^1(\tilde{\mathbf{x}}_d^2)^\top] = \mathbf{0}$ while $\hat{\mathbf{x}}_d^1$ and $\hat{\mathbf{x}}_d^2$ are correlated to an unknown degree. A simple example is the fusion of two estimates after local measurement updates, which allows us to exploit the conditional independence of the measurements [73]. In split CI, the joint form (10) can again be exploited to find the bound

$$\mathbb{E}\left[\tilde{\mathbf{x}}^J(\tilde{\mathbf{x}}^J)^\top\right] = \begin{bmatrix}\mathbf{P}_e^1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_e^2\end{bmatrix} + \begin{bmatrix}\mathbf{P}_d^1 & \mathbf{P}_d^{12} \\ \mathbf{P}_d^{21} & \mathbf{P}_d^2\end{bmatrix}$$
$$\preceq \begin{bmatrix}\mathbf{P}_e^1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_e^2\end{bmatrix} + \begin{bmatrix}\frac{1}{\omega}\mathbf{P}_d^1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{1-\omega}\mathbf{P}_d^2\end{bmatrix}$$

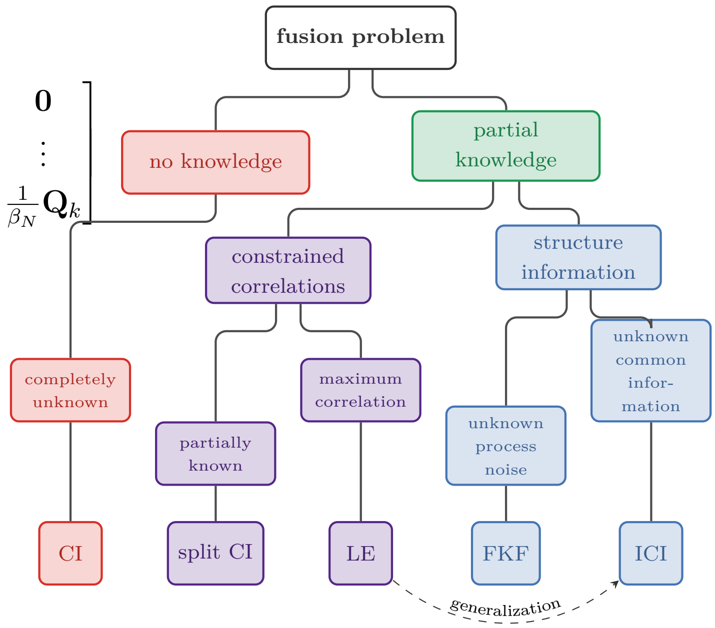


Figure 12: Coarse classification of alternative approaches to CI that require different knowledge about the fusion problem. If correlations are completely unknown, CI is optimal. If correlations can be constrained, split CI or the largest ellipsoid (LE) method can be considered. Information about the sources of correlations can be utilized in the federated Kalman filter (FKF) and inverse covariance intersection (ICI). The former handles common process noise; the latter is tailored to unknown common information.

with $\omega \in [0,1]$, $\mathbf{P}_e^i = \mathbb{E}[\tilde{\mathbf{x}}_e^i(\tilde{\mathbf{x}}_e^i)^\top]$ and $\mathbf{P}_d^i = \mathbb{E}[\tilde{\mathbf{x}}_d^i(\tilde{\mathbf{x}}_d^i)^\top]$, and $\mathbf{P}_d^{12} = (\mathbf{P}_d^{21})^\top = \mathbb{E}[\tilde{\mathbf{x}}_d^1(\tilde{\mathbf{x}}_d^2)^\top]$. Hence, the bound is only applied to dependent parts.

Split CI indirectly assumes a bound on the maximum possible cross-covariance matrix $\mathbf{P}^{12}$ between the estimates $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$, which is characterized through

$$(\mathbf{P}_d^{12})^\top(\mathbf{P}_d^1)^{-1}\mathbf{P}_d^{12} \preceq \mathbf{P}_d^2 \,.$$

The authors of [74] and [75] independently have proposed to utilize a similar form of a bound

$$(\mathbf{P}^{12} - \mathbf{D})^\mathsf{T}(\mathbf{P}^1)^{-1}(\mathbf{P}^{12} - \mathbf{D}) \preceq r_{\max}^2 \mathbf{P}^2 \ .$$

for $\mathbf{P}^{12}$ with $r_{\max} \in (0,1]$. This relationship specifies how much the unknown $\mathbf{P}^{12}$ can differ from a known $\mathbf{D}$. If such $\mathbf{D}$ and $r_{\max}$ are known, the joint cross-covariance matrix (10) can then be bounded by

$$\mathbf{P}^J \preceq \begin{bmatrix} (1+\alpha \cdot r_{\max})\mathbf{P}^1 & \mathbf{D} \\ \mathbf{D}^\mathsf{T} & (1+\frac{r_{\max}}{\alpha})\mathbf{P}^2 \end{bmatrix}$$

with $\alpha = \frac{1}{\omega} - 1$, which can then be used in (13) for conservative fusion. Specific bounds and limiting cases are studied further in [76].

While split CI uses an additive splitting of the estimates, [77] studies a splitting along the state components. For instance, if the state estimates have the form $\hat{\mathbf{x}}^1 = [(\hat{\mathbf{x}}_a^1)^\mathsf{T}\ (\hat{\mathbf{x}}_b^1)^\mathsf{T}]^\mathsf{T}$ and $\hat{\mathbf{x}}^2 = [(\hat{\mathbf{x}}_c^2)^\mathsf{T}\ (\hat{\mathbf{x}}_d^2)^\mathsf{T}]^\mathsf{T}$, where only $\hat{\mathbf{x}}_b^1$ and $\hat{\mathbf{x}}_d^2$ have unknown correlations and all other correlations are known, the bound

$$\mathbf{P}^J = \begin{bmatrix} \mathbf{P}_{aa}^1 & \mathbf{P}_{ab}^1 & \mathbf{P}_{ac}^{12} & \mathbf{P}_{ad}^{12} \\ \mathbf{P}_{ba}^1 & \mathbf{P}_{bb}^1 & \mathbf{P}_{bc}^{12} & \mathbf{P}_{bd}^{12} \\ \mathbf{P}_{ca}^{21} & \mathbf{P}_{cb}^1 & \mathbf{P}_{cc}^2 & \mathbf{P}_{cd}^2 \\ \mathbf{P}_{da}^{21} & \mathbf{P}_{db}^{21} & \mathbf{P}_{dc}^2 & \mathbf{P}_{dd}^2 \end{bmatrix} \preceq \begin{bmatrix} \mathbf{P}_{aa}^1 & \mathbf{P}_{ab}^1 & \mathbf{P}_{aa}^{12} & \mathbf{P}_{ab}^{12} \\ \mathbf{P}_{ba}^1 & \frac{1}{\omega}\mathbf{P}_{bb}^1 & \mathbf{P}_{bc}^{12} & \mathbf{0} \\ \mathbf{P}_{ca}^{21} & \mathbf{P}_{cb}^{21} & \mathbf{P}_{cc}^2 & \mathbf{P}_{cd}^2 \\ \mathbf{P}_{da}^{21} & \mathbf{0} & \mathbf{P}_{dc}^2 & \frac{1}{1-\omega}\mathbf{P}_{dd}^2 \end{bmatrix}$$

highlighted in red can be utilized. In [78] it is noticed that this bound is overly conservative, which is due to the fact that the known correlations of the other state components limit the correlations between $\hat{\mathbf{x}}_b^1$ and $\hat{\mathbf{x}}_d^2$. The same work demonstrates how to find a tight bound that can be exploited for fusion.

A related concept is the factorized covariance intersection (FCI, [79]–[81]) method that exploits a joint diagonalization [79], [82] for reducing the conservativeness of traditional CI. To illustrate the general concept, we again assume estimates of the form $\hat{\mathbf{x}}^1 = [(\hat{\mathbf{x}}_a^1)^\mathsf{T}\ (\hat{\mathbf{x}}_b^1)^\mathsf{T}]^\mathsf{T}$ and $\hat{\mathbf{x}}^2 = [(\hat{\mathbf{x}}_a^2)^\mathsf{T}\ (\hat{\mathbf{x}}_b^2)^\mathsf{T}]^\mathsf{T}$. After a joint diagonalization, it is assumed that they have the joint covariance matrix

$$\mathbf{P}^J = \begin{bmatrix} \mathbf{P}^1 & \mathbf{P}^{12} \\ \mathbf{P}^{21} & \mathbf{P}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{P}_a^1 & \mathbf{0} & \mathbf{P}_a^{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_b^1 & \mathbf{0} & \mathbf{P}_b^{12} \\ \mathbf{P}_a^{21} & \mathbf{0} & \mathbf{P}_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_b^{21} & \mathbf{0} & \mathbf{P}_b^2 \end{bmatrix} . \tag{25}$$

If this factorization holds, it is possible to handle the a (red) and b (blue) block components separately. An estimate $\hat{\mathbf{x}}$ can then be computed using FCI as

$$\hat{\mathbf{x}} = \begin{bmatrix} \hat{\mathbf{x}}_a \\ \hat{\mathbf{x}}_b \end{bmatrix} = \begin{bmatrix} \mathbf{P}_a^{\omega_a}\left(\omega_a(\mathbf{P}_a^1)^{-1}\hat{\mathbf{x}}_a^1 + (1-\omega_a)(\mathbf{P}_a^2)^{-1}\hat{\mathbf{x}}_a^2\right) \\ \mathbf{P}_b^{\omega_b}\left(\omega_b(\mathbf{P}_b^1)^{-1}\hat{\mathbf{x}}_b^1 + (1-\omega_b)(\mathbf{P}_b^2)^{-1}\hat{\mathbf{x}}_b^2\right) \end{bmatrix} ,$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_a^{\omega_a} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_b^{\omega_b} \end{bmatrix}$$

with

$$\mathbf{P}_a^{\omega_a} = \left(\omega_a(\mathbf{P}_a^1)^{-1} + (1-\omega_a)(\mathbf{P}_a^2)^{-1}\right)^{-1} ,$$
$$\mathbf{P}_b^{\omega_b} = \left(\omega_b(\mathbf{P}_b^1)^{-1} + (1-\omega_b)(\mathbf{P}_b^2)^{-1}\right)^{-1} ,$$

and $\omega_a, \omega_b \in [0,1]$ are found by optimizing $\mathbf{P}_a^{\omega_a}$ and $\mathbf{P}_b^{\omega_b}$ separately. As noted in [81] it is sufficient to have the block structure in (25) after a joint transformation of $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$.

Largest Ellipsoid Methods

While CI focuses on the bound of the intersection as shown in Figure 4, a famous class of conservative fusion algorithms studies the largest ellipsoid inside the intersection. This idea was initially proposed in [83] and has later on been given different names. In [84] it is named internal ellipsoid approximation and in the same paper it is pointed out that the fused estimate mean has not been derived correctly in the initial work. The authors of [85], [86] use the name ellipsoidal intersection and, in [87], [88], it goes by the name safe fusion. While they all introduce systematic ways to derive the fused estimates, there are some differences between how the state estimates are computed, which are not further elucidated. The LE method formulation of [82] is used here.

An intuitive description of LE is obtained through a joint diagonalization of the estimates. Let

$$\begin{aligned} \hat{\mathbf{z}}^1 &= \mathbf{T}\hat{\mathbf{x}}^1, & (\mathbf{D}^1)^{-1} &= \mathbf{T}(\mathbf{P}^1)^{-1}\mathbf{T}^\top, \\ \hat{\mathbf{z}}^2 &= \mathbf{T}\hat{\mathbf{x}}^2, & (\mathbf{D}^2)^{-1} &= \mathbf{T}(\mathbf{P}^2)^{-1}\mathbf{T}^\top \end{aligned}$$

be the estimates after applying a transformation $\mathbf{T}$ that yields diagonal $\mathbf{D}^1$ and $\mathbf{D}^2$. In essence, $\mathbf{T}$ scales and rotates both ellipsoids $\mathcal{E}(\mathbf{P}^1)$ and $\mathcal{E}(\mathbf{P}^2)$ to be axis aligned. Such a diagonalization is, for instance, desribed in [86]. To fuse both estimates, LE selects the elements with the smallest covariance entries, that is,

$$[\hat{\mathbf{z}}^{\mathrm{LE}}]_i = \begin{cases} [\hat{\mathbf{z}}^1]_i \,, & \text{if } [\mathbf{D}^1]_{ii} < [\mathbf{D}^2]_{ii} \,, \\ [\hat{\mathbf{z}}^2]_i \,, & \text{if } [\mathbf{D}^1]_{ii} > [\mathbf{D}^2]_{ii} \,, \\ \frac{1}{2}\Big([\hat{\mathbf{z}}^1]_i + [\hat{\mathbf{z}}^2]_i\Big) \,, & \text{if } [\mathbf{D}^1]_{ii} = [\mathbf{D}^2]_{ii} \end{cases}$$

$$[\mathbf{D}^{\mathrm{LE}}]_{ii} = \min\Big\{[\mathbf{D}^1]_{ii}, [\mathbf{D}^2]_{ii}\Big\} \ .$$

where $[\mathbf{D}]_{ii}$ is the ith diagonal entry of matrix $\mathbf{D}$ and $[\mathbf{z}]_i$ the ith entry of vector $\mathbf{z}$. The fusion result is finally obtained by the back transformations $\hat{\mathbf{x}}^{\mathrm{LE}} = \mathbf{T}^{-1}\hat{\mathbf{z}}^{\mathrm{LE}}$ and $\mathbf{P}^{\mathrm{LE}} = \mathbf{T}^{-1}\mathbf{D}^{\mathrm{LE}}\mathbf{T}^{-\top}$. The equations show that LE computes the maximum-volume ellipsoid, also called Löwner–John ellipsoid, that fits into the intersection $\mathcal{E}(\mathbf{P}^1) \cap \mathcal{E}(\mathbf{P}^2)$.

The LE method can also be expressed in terms of the assumption that the two estimates $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ share common information $\hat{\boldsymbol{\gamma}}$ with covariance matrix $\boldsymbol{\Gamma}$ as in (15), that is,

$$\begin{aligned} \hat{\mathbf{x}}^1 &= \mathbf{P}^1\Big((\mathbf{P}^1_e)^{-1}\hat{\mathbf{x}}^1_e + \boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\gamma}}\Big) \ , \\ \hat{\mathbf{x}}^2 &= \mathbf{P}^2\Big((\mathbf{P}^2_e)^{-1}\hat{\mathbf{x}}^2_e + \boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\gamma}}\Big) \end{aligned} \tag{26}$$

with

$$\begin{aligned} \mathbf{P}^1 &= \Big((\mathbf{P}^1_e)^{-1} + \boldsymbol{\Gamma}^{-1}\Big)^{-1} , \\ \mathbf{P}^2 &= \Big((\mathbf{P}^2_e)^{-1} + \boldsymbol{\Gamma}^{-1}\Big)^{-1} , \end{aligned} \tag{27}$$

where $\hat{\mathbf{x}}^1_e$, $\hat{\mathbf{x}}^2_e$, and $\hat{\boldsymbol{\gamma}}$ are mutually uncorrelated. The estimates $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ are correlated through the common information such that $\mathbf{P}^{12} = \mathbb{E}[\tilde{\mathbf{x}}^1(\tilde{\mathbf{x}}^2)^\top] = \mathbf{P}^1\boldsymbol{\Gamma}^{-1}\mathbf{P}^2$. Optimal fusion, in this case, implies

$$\begin{aligned} \hat{\mathbf{x}}^{\mathrm{fus}} &= \mathbf{P}\Big((\mathbf{P}^1)^{-1}\hat{\mathbf{x}}^1 + (\mathbf{P}^2)^{-1}\hat{\mathbf{x}}^2 - \boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\gamma}}\Big) \\ &= \mathbf{P}\Big((\mathbf{P}^1_e)^{-1}\hat{\mathbf{x}}^1_e + (\mathbf{P}^2_e)^{-1}\hat{\mathbf{x}}^2_e + \boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\gamma}}\Big) \ , \end{aligned} \tag{28a}$$

$$\begin{aligned} \mathbf{P}^{\mathrm{fus}} &= \Big((\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1} - \boldsymbol{\Gamma}^{-1}\Big)^{-1} \\ &= \Big((\mathbf{P}^1_e)^{-1} + (\mathbf{P}^2_e)^{-1} + \boldsymbol{\Gamma}^{-1}\Big)^{-1} , \end{aligned} \tag{28b}$$

meaning that $\hat{\mathbf{x}}^1_e$, $\hat{\mathbf{x}}^2_e$, and $\hat{\boldsymbol{\gamma}}$ are fused optimally. Since $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{\Gamma}$ are unknown, the LE method selects $\hat{\boldsymbol{\gamma}}$ with the largest information matrix $\boldsymbol{\Gamma}^{-1}$ in terms of volume. It hence subtracts the maximum possible common information.

The LE method's effectiveness is shown in empirical case studies [85] and [87]. A comparison with CI is conducted in [89]. In [68], it is shown under which assumptions on $\mathbf{P}^{12}$ LE is an optimal CLUE. However, the LE method provides conservative estimates only in certain rather restrictive cases [90]. The reason is that LE refers only to the largest possible common information in terms of the volume of $\mathcal{E}(\mathbf{P}^{\text{LE}})$, that is, the determinant of $\mathbf{P}^{\text{LE}}$, but there can be instances of common information such that $\mathcal{E}(\mathbf{P}^{\text{LE}})$ will be exceeded in some directions.

Inverse Covariance Intersection

Inverse covariance intersection (ICI, [90]) addresses the problem of unknown common information by providing a conservative bound on all possible instances of common information and, thus, can also handle those that are not captured by the LE method. ICI refers to the same decomposition as the LE method in (26) and (27). ICI exploits the bound

$$\mathbf{\Gamma}^{-1} \preceq \left(\omega \mathbf{P}^1 + (1-\omega)\mathbf{P}^2\right)^{-1} \qquad (29)$$

on all possible $\mathbf{\Gamma}^{-1}$ with $\omega \in [0,1]$. Note the resemblance to $\mathbf{P}^{\text{CI}}$ in (5b) of CI. The difference lies in which domain this bounding takes place. In case of CI, it is done in the covariance domain and, in case of ICI, it is done in the information domain. Fusion of $\hat{\mathbf{x}}^1$ and $\hat{\mathbf{x}}^2$ using ICI yields

$$\hat{\mathbf{x}}^{\text{ICI}} = \mathbf{P}^{\text{ICI}}\Big((\mathbf{P}^1)^{-1}\,\hat{\mathbf{x}}^1 + (\mathbf{P}^2)^{-1}\,\hat{\mathbf{x}}^2 - (\mathbf{\Gamma}^{\text{ICI}})^{-1}\,\hat{\boldsymbol{\gamma}}^{\text{ICI}}\Big),$$

$$\mathbf{P}^{\text{ICI}} = \Big((\mathbf{P}^1)^{-1} + (\mathbf{P}^2)^{-1} - (\mathbf{\Gamma}^{\text{ICI}})^{-1}\Big)^{-1}$$

with the bound on common information

$$\hat{\boldsymbol{\gamma}}^{\text{ICI}} = \omega\,\hat{\mathbf{x}}^1 + (1-\omega)\,\hat{\mathbf{x}}^2 \ ,$$
$$\mathbf{\Gamma}^{\text{ICI}} = \omega \mathbf{P}^1 + (1-\omega)\mathbf{P}^2 \ ,$$

where ω is computed by optimizing a loss function $J(\mathbf{P})$ like the trace of $\mathbf{P}$. As it can be seen by comparison to (28b), ICI uses the bound (29) in place of the unknown $\mathbf{\Gamma}$.

ICI guarantees conservative estimates given that the decomposition in (26)–(27) is satisfied. Moreover, it is shown that ICI provides a tight bound on $\mathbf{P}$ and, hence, is an MSE optimal conservative linear estimator, which means that it minimizes the MSE among all conservative linear estimators, under these assumptions. In [91], a generalization of the decomposition in (26)–(27) is derived. Different aspects of ICI and the fusion of multiple estimates are studied in [92]. An alternative parameterization of the fusion gains is provided in [93]. An empirical study of ICI and comparison with CI is provided by [88]. Formulas for a batchwise fusion of multiple estimates using ICI are derived in [94], and [95] studies iterative solutions. ICI is an optimal CLUE given that the common information decomposition [68].

Discussion and Other Alternatives

Figure 12 depicts that each method described utilizes different assumptions regarding the correlation structure. Split CI and the LE method can exploit partial knowledge or constraints on the joint covariance matrix. This knowledge is only accessible in specific cases, such as when the state comprises multiple objects and only some of the correlations are unknown while the correlations between the other object states are available. The LE method assumes a maximum possible common information shared by two estimates and hence addresses the problem of double counting in (15). Its generalization ICI bounds all possible instances of common information. Both methods struggle when also common process noise is present and blends with the common information. FKF takes the opposite direction and is designed explicitly for common process noise. Therefore, it struggles with common information. However, the inseparability of common information and process noise over time does not prohibit using these methods. For instance, [91] shows that ICI can also handle common process noise to some degree.

The discussed methods do not form a complete list, and also alternatives have been explored. In [96], a min-max formulation is suggested where the MSE is minimized for the maximum possible cross-correlations. A game-theoretic approach is considered in [97] in terms of a two-player game, where one player tries to minimize the MSE while the other tunes the correlations to maximize the MSE. A full maximum cross-covariance is computed in [98]. These approaches have in common that they try to represent the maximum possible correlations. However, as the maximum possible correlations typically are not unique these approaches cannot in general guarantee conservativeness. In [39], a simple parameterization of fusion results given a bound on the correlations is derived, which ranges from the independent case to full CI. Yet another approach is the optimization-based approach suggested in [68]. Using robust optimization it is possible to solve for an optimal CLUE in general cases by explicitly defining the partial knowledge about the cross-correlations as uncertain constraints.

Many of the studied methods have also counterparts for nonlinear fusion problems. The implications and challenges in nonlinear settings are briefly discussed in *Sidebar: Conservative Non-Gaussian Fusion*.

Conclusions and Outlook

With the increasing deployment of decentralized data fusion (DDF) systems, a significant portion of the research focus within distributed and decentralized estimation is being devoted to covariance intersection (CI) and related algorithms. CI exhibits the remarkable feature of producing conservative and, thus, reliable estimates even though dependencies among estimates to be fused are unknown. This property leverages distributed and, in particular, DDF applications like mobile sensor networks. In this article, we have discussed the underlying theory and essential properties such as optimality and conservativeness of CI, and illustrated the considered estimation problem using different examples. CI has sparked several research questions addressing computational aspects, the exploitation of additional information, and the reduction of over-conservativeness. While the complexity and richness of CI-related research have steadily increased, this article hopefully provides easy access to CI for both the practitioner and researcher. Undoubtedly, the appeal of CI lies in that it is easy to implement and robustly handles any possible cross-correlations between estimates to be fused. The following and concluding discussion summarizes past and present development giving us an opportunity for a future outlook.

Past to Present

Past and present research endeavors related to CI, its alternatives, and its applications primarily draw their motivation from three key aspects: *modularity*, *scalability*, and *optimality*.

Modularity. Developing technology with a modular design will be even more crucial in the future. Without modularity, the adverse effects of legacy systems and system complexity will eventually become overwhelming, both for the design engineer and any other personnel who maintains or operates the system at hand. To this end, algorithms such as CI have a clear advantage. CI is simple to implement yet powerful enough to allow for an overall modular design as it robustly handles cross-correlations between estimates. This opens CI up for a vast amount of DDF applications.

Scalability. With the advent of affordable, integrated, and connected sensor technology, DDF systems have to handle a huge number of measurements from sensors that offer high availability and are abundant in number. Moreover, with modern sensor systems, new algorithms, and more computational power, it is possible to measure new quantities, generate more measurements and estimate additional state components, which leads to potentially high-dimensional state estimates. Such estimates can be a challenge even for small sensor networks, and these estimates often scale poorly in large sensor networks. In such configurations CI is an attractive algorithm. CI does not require exchanging anything more than a vector and a matrix. In addition, CI requires neither a mechanism to keep track of correlations nor a particular communication or filtering scheme to be deployed.

Optimality. CI is an optimal conservative fusion algorithm given that the cross-correlations are completely unknown. Optimality means that no other conservative solution with a smaller covariance matrix is attainable. Yet, the conservativeness is a weak spot of CI that could lead to fusion results with a relatively marginal information gain. When partial knowledge about the cross-correlation structure is available, CI becomes over-conservative, and less conservative fusion results with a smaller covariance matrix are possible. For this reason, a lot of work related to CI has been dedicated to alternatives and derivatives that more efficiently exploit available partial knowledge.

Some of the key milestones of the research related to CI conducted so far are summarized in Figure 13. During the first time after the invention of CI in (the year of) 1997, a primary focus resided on related linear estimation methods during which alternatives such as split CI and the largest ellipsoid (LE) method were derived. Since, at that time, computational resources were more restricted than today, much of the focus was on cheap and simple solutions, for example, fast CI. Remarkably, the interest and necessity in theoretical aspects of CI is still unbroken. This is also true regarding linear fusion methods where inverse covariance intersection (ICI) was developed just a few years ago. Recently, a generalized methodology defining conservative linear unbiased estimators (CLUE) has been suggested to capture estimation problems under partially known covariances. These developments demonstrate the current trend to exploit partial knowledge to compute fused estimates more tightly.

Concerning applications, object tracking and simultaneous localization and mapping (SLAM) are the two main application areas. This is quite natural since decentralized sensor networks are highly relevant in these applications. In the last decade, CI has been implemented successfully in the field of robotics. Current developments indicate that CI and its derivatives will become important in autonomous driving and vehicle-to-everything (V2X) networks, motivated by the huge amount of data transfer and the high level of modularity in these configurations.

Present to Future

CI shows significant potential to remain a key topic of future research in the upcoming 25 years. From a today's perspective, promising and prospective directions concerning open problems are *machine learning*, *exploiting structure*, *nonlinear fusion*, and *data efficiency*.

Machine Learning. The extraordinary spread of data-driven techniques, such as machine learning methods, follows from their ability to solve very complex problems in diverse scientific fields. Their potential in DDF is, for instance, emphasized in [99]. Two of many subtopics within machine learning which are only partially covered are (i) uncertainty assessment and (ii) how to combine classification and/or estimation results from different sources, both of which depend on correlations between parameters. Since CI handles dependencies robustly, it is a poten-

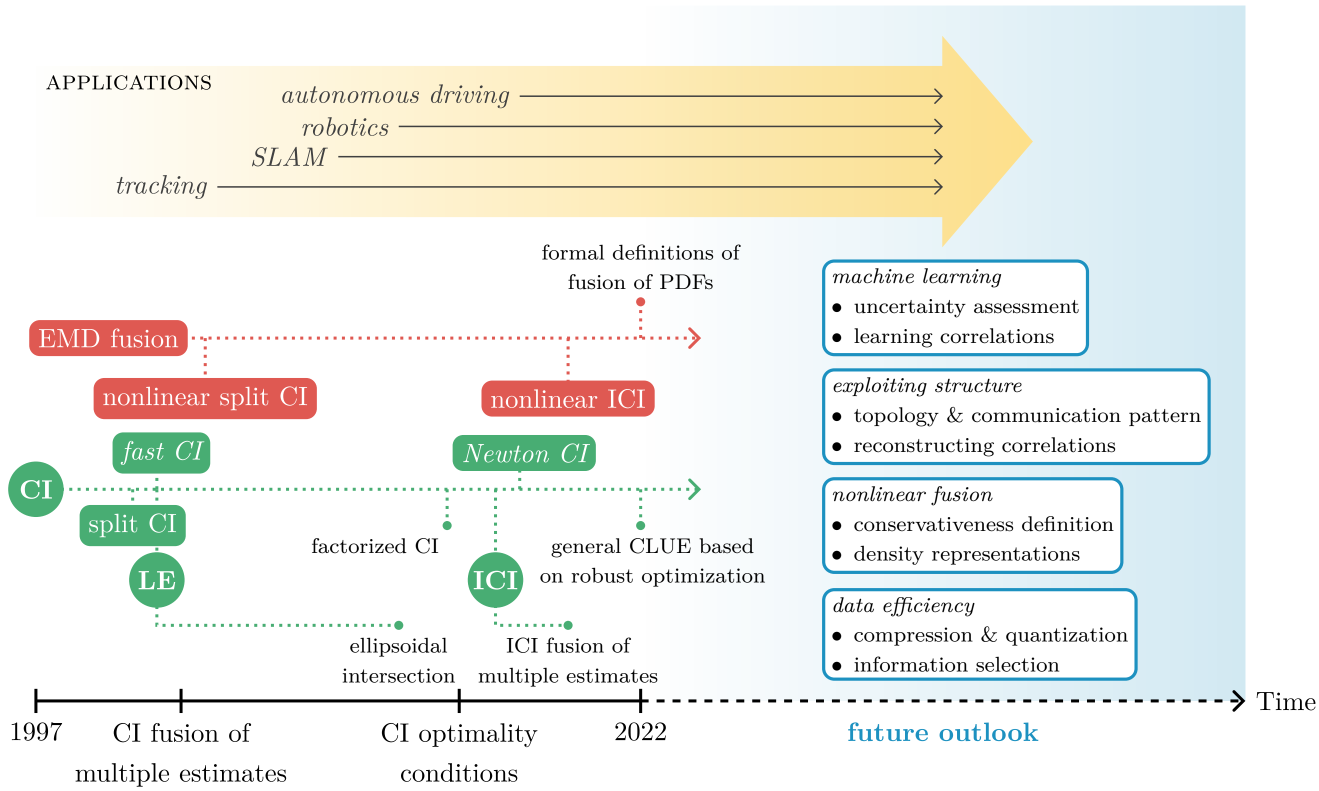


Figure 13: Summary of conducted research related to CI until now. Much of the attention, both in the present and past, have been drawn to developing alternatives to CI that more efficiently exploits the available information. Generalizations of CI and its derivatives for the handling of arbitrary probability densities are also ongoing work. The main application areas have been relatively constant, where object tracking and SLAM still are the most studied ones. A summary of the future outlook is provided in the right part.

tial key ingredient for correct uncertainty assessment and merging of information derived using machine learning. Moreover, machine learning allows us to partially learn and predict covariance structures and dependencies present in different scenarios such that less conservative alternatives of CI are applicable. For example, [100] learns bounds on possible correlations, and [101] addresses the computational costs using machine learning.

Exploiting Structure. Several alternatives to CI rely on specific correlation structures, which are either assumed based on the DDF problem at hand or reconstructed under certain assumptions. These correlation structures can be derived by extracting information from the topology and communication structure of the DDF network, for which factor graphs prove to be useful [102]. Also, a sample-based reconstruction of correlation structure can be an option [103]. As pointed out above, an interesting direction is how to use machine learning to partially learn the correlation structure. If this is accomplished, it will be possible to obtain tighter fusion results. Another direction can be the relaxation of the conservativeness constraint by replacing the covariance bound with a probabilistic constraint. For such an approach, the stochastic optimization paradigm has a clear potential similar to how robust optimization is utilized in the general CLUE problem studied in [68].

Nonlinear Fusion. As stated in *Sidebar: Conservative Non-Gaussian Fusion* and the references therein, applying CI to nonlinear fusion problems remains challenging. Nonlinear versions of CI and its derivatives have been developed, but these are tailored to specific problems where Gaussian estimates cannot be obtained, and more general probability densities need to be considered. An open problem is how to derive a nonlinear counterpart of conservative fusion in general settings. Still, such nonlinear fusion problems lack a proper definition of conservativeness and, thus, nonlinear fusion algorithms cannot provide the same guarantees yet as CI for the linear case. DDF with non-Gaussian distributions leads to further problems as they typically require more parameters than a Gaussian density representation, and computational aspects may become a significant concern.

Data Efficiency. Besides nonlinear settings, data volume will increase with advances in sensor technology and connectivity that, for instance, allow for sharing tracking results and maps in V2X applications. At the same time, data efficiency becomes a dominant aspect in long-range wide-area networks and battery-driven Internet-of-Things applications, and it will become important to optimize the data transfer, for instance, by compressing data. Different aspects of this data compression problem are investigated in, for instance, [49], [51], [53], [54]. CI can significantly contribute to robustness when data needs to be compressed, and cross-covariance information is approximated or ignored. Another technique is selecting the most informative data when high-dimensional state representations need to be exchanged. Such preprocessing will alter the possible correlations that can also be treated by CI [52].

Correlations Still Unknown?

CI is a versatile method that fuses estimates irrespective of any cross-correlations. Yet, this article shows that fusion under unknown dependencies is not solved. CI can limit the information gain, and one may not accept too conservative fusion results, so a reasonable trade-off between informativeness and conservativeness is to be found. Hence, CI will form a toolset with its derivatives like factorized CI, ICI, and the CLUE approach and will flexibly adapt to the fusion problem at hand. Future research will extend this toolset. Besides the challenging field of nonlinear fusion, conservative fusion has to keep pace with the current trends, where model-based and data-driven sensor fusion come together. The CI toolset will undoubtedly be an indispensable backbone of future DDF systems.

Acknowledgements

This work has been supported by the Industry Excellence Center LINK-SIC funded by The Swedish Governmental Agency for Innovation Systems (VINNOVA) and Saab AB.

We would like to thank the reviewers for taking the necessary time and effort to review the manuscript. We especially appreciate your valuable suggestions to enhance the historical background, which have made this article more complete.

References

[1] H. Qiu, M. Qiu, Z. Lu, and G. Memmi, "An efficient key distribution system for data fusion in V2X heterogeneous networks," *Information Fusion*, vol. 50, pp. 212–220, 2019.

[2] V. Shin, V. Hamdipoor, and Y. Kim, "Reduced-order multisensory fusion estimation with application to object tracking," *IET Signal Processing*, vol. 16, no. 4, pp. 463–478, Apr. 2022.

[3] S. J. Julier and J. K. Uhlmann, "Simultaneous localisation and map building using split covariance intersection," in *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Maui, HI, USA, Nov. 2001.

[4] ——, "Using covariance intersection for SLAM," *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 3–20, Jan. 2007.

[5] B. Noack, S. J. Julier, and U. D. Hanebeck, "Treatment of biased and dependent sensor data in graph-based SLAM," in *Proceedings of the 18th IEEE International Conference on Information Fusion (FUSION)*, Washington D.C., USA, Jul. 2015.

[6] Z. Zhu and C. N. Taylor, "Conservative uncertainty estimation in map-based vision-aided navigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 2, pp. 941–949, Apr. 2017.

[7] T. R. Wanasinghe, G. K. I. Mann, and R. G. Gosine, "Decentralized cooperative localization approach for autonomous multirobot systems," *Journal of Robotics*, 2016.

[8] J. Klingner, N. Ahmed, and N. Correll, "Fault-tolerant covariance intersection for localizing robot swarms," *Robotics and Autonomous Systems*, vol. 122, p. 103306, 2019.

[9] J. Lai, Y. Zhou, J. Lin, Y. Cong, and J. Yang, "Cooperative localization based on efficient covariance intersection," *IEEE Communications Letters*, vol. 23, no. 5, pp. 871–874, 2019.

[10] S. J. Julier and J. J. LaViola Jr., "An empirical study into the robustness of split covariance addition (SCA) for human motion tracking," in *Proceedings of the 2004 American Control Conference (ACC)*, Boston, MA, USA, Jun. 2004.

[11] U. D. Hanebeck and J. Horn, "Data validation in the presence of imprecisely known correlations," in *Proceedings of the 2003 European Control Con-*

ference (ECC), Cambridge, UK, Sep. 2003.

[12] J. Zhang, G. Scebba, and W. Karlen, "Covariance intersection to improve the robustness of the photoplethysmogram derived respiratory rate," in *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Montréal, Canada: IEEE, Jul. 2020.

[13] 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.

[14] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proceedings of the 1997 American Control Conference (ACC)*, vol. 4, Albuquerque, NM, USA, Jun. 1997, pp. 2369–2373.

[15] S. Grime and H. F. Durrant-Whyte, "Data fusion in decentralized sensor networks," *Control Engineering Practice*, vol. 2, no. 5, pp. 849–863, Oct. 1994.

[16] D. Smith and S. Singh, "Approaches to multisensor data fusion in target tracking: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 12, pp. 1696–1710, Oct. 2006.

[17] M. E. Liggins II, 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.

[18] D. L. Hall, C.-Y. Chong, J. Llinas, and M. E. Liggins II, Eds., *Distributed Data Fusion for Network-Centric Operations*, ser. The Electrical Engineering and Applied Signal Processing Series. CRC Press, 2013.

[19] A. G. O. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*. Boca Raton, Florida, USA: CRC Press, Inc., 1998.

[20] J. K. Uhlmann, "General data fusion for estimates with unknown cross covariances," in *Proc. of SPIE – Signal Processing, Sensor Fusion, and Target Recognition V*, vol. 2755, Orlando, FL, USA, Jun. 1996.

[21] L. Chen, P. O. Arambel, and R. K. Mehra, "Fusion under unknown correlation – covariance intersection as a special case," in *Proceedings of the 5th IEEE International Conference on Information Fusion (FUSION)*, Annapolis, MD, USA, Jul. 2002.

[22] U. Orguner, "Approximate analytical solutions for the weight optimization problems of CI and ICI," in *Proceedings of the 2017 IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, Oct. 2017, pp. 1–6.

[23] X.-R. Li, Y. Zhu, J. Wang, and C. Han, "Optimal linear estimation fusion—part i: Unified fusion rule," *IEEE*

Transactions on Information Theory, vol. 49, no. 9, pp. 2192–2208, Sep. 2003.

[24] J. R. Carpenter and R. H. Bishop, "Navigation filter estimate fusion for enhanced spacecraft rendezvous," *Journal of Guidance, Control, and Dynamics*, vol. 20, no. 2, pp. 338–345, 1997.

[25] B. Noack, J. Sijs, and U. D. Hanebeck, "Fusion strategies for unequal state vectors in distributed Kalman filtering," in *Proceedings of the 19th IFAC World Congress*, Cape Town, South Africa, Aug. 2014.

[26] J. R. A. Klemets and M. Hovd, "Hierarchical decentralized state estimation with unknown correlation for multiple and partially overlapping state vectors," in *Proceedings of the 2018 IEEE Conference on Control Technology and Applications (CCTA)*, Copenhagen, Denmark, Jul. 2018.

[27] I. Loefgren, N. Ahmed, E. Frew, C. Heckman, and S. Humbert, "Scalable event-triggered data fusion for autonomous cooperative swarm localization," in *Proceedings of the 22nd International Conference on Information Fusion (FUSION)*. Ottawa, Canada: IEEE, Jul. 2019.

[28] J. K. Uhlmann, "Covariance consistency methods for fault-tolerant distributed data fusion," *Information Fusion*, vol. 4, no. 3, pp. 201–215, 2003.

[29] L. Chen, P. O. Arambel, and R. K. Mehra, "Fusion under unknown correlation: Covariance intersection revisited," *IEEE Transactions on Automatic Control*, vol. 47, no. 11, pp. 1879–1882, Nov. 2002.

[30] B. Noack, *State Estimation for Distributed Systems with Stochastic and Set-membership Uncertainties*, ser. Karlsruhe Series on Intelligent Sensor-Actuator-Systems 14. Karlsruhe, Germany: KIT Scientific Publishing, 2013.

[31] B. Noack, J. Sijs, M. Reinhardt, and U. D. Hanebeck, "Treatment of dependent information in multisensor Kalman filtering and data fusion," in *Multisensor Data Fusion: From Algorithms and Architectural Design to Applications*, H. Fourati, Ed. CRC Press, Aug. 2015, pp. 169–192.

[32] S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion Kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, Jun. 2004.

[33] Y. Bar-Shalom and L. Campo, "On the track-to-track correlation problem," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 571–572, Apr. 1981.

[34] M. A. Bakr and S. Lee, "Distributed multisensor data fusion under unknown correlation and data inconsistency," *Sensors*, vol. 17, no. 11, p. 2472, Oct. 2017.

[35] S. J. Julier, "Fusion without independence (keynote abstract)," in *Proceedings of the IET Seminar on Tracking and Data Fusion: Algorithms and Applications*, Birmingham, UK, Apr.

2008, pp. 1–5.

[36] L. Yan, X.-R. Li, Y. Xia, and M. Fu, "Optimal sequential and distributed fusion for state estimation in cross-correlated noise," *Automatica*, vol. 49, no. 12, pp. 3607–3612, Dec. 2013.

[37] A. Jindal and K. Psounis, "Modeling spatially-correlated sensor network data," in *Proceedings of the 1st IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, Santa Clara, CA, USA, Jan. 2004, pp. 162–171.

[38] M. Reinhardt, B. Noack, P. O. Arambel, and U. D. Hanebeck, "Minimum covariance bounds for the fusion under unknown correlations," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1210–1214, Sep. 2015.

[39] Z. Wu, Q. Cai, and M. Fu, "Covariance intersection for partially correlated random vectors," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 619–629, Mar. 2018.

[40] K.-C. Chang, C.-Y. Chong, and S. Mori, "Analytical and computational evaluation of scalable distributed fusion algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 4, pp. 2022–2034, Oct. 2010.

[41] M. B. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Proceedings of the 5th IEEE International Conference on Information Fusion (FUSION)*, Annapolis, MD, USA, Jul. 2002.

[42] W. Niehsen, "Information fusion based on fast covariance intersection filtering," in *Proceedings of the 5th IEEE International Conference on Information Fusion (FUSION)*, Annapolis, MD, USA, Jul. 2002.

[43] J. R. Carpenter and R. H. Bishop, "Estimate fusion for lunar rendezvous," in *Proceedings of the Guidance, Navigation and Control Conference*, Monterey, CA, USA, Aug. 1993, pp. 1–10.

[44] D. Fränken and A. Hüpper, "Improved fast covariance intersection for distributed data fusion," in *Proceedings of the 8th IEEE International Conference on Information Fusion (FUSION)*, Philadelphia, PA, USA, Jul. 2005.

[45] Y. Wang and X.-R. Li, "A fast and fault-tolerant convex combination fusion algorithm under unknown cross-correlation," in *Proceedings of 12th IEEE International Conference on Information Fusion (FUSION)*, Seattle, WA, USA, Jul. 2009.

[46] ——, "Distributed estimation fusion with unavailable cross-correlation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 259–278, Jan. 2012.

[47] M. Ristic, B. Noack, and U. D. Hanebeck, "Secure fast covariance intersection using partially homomorphic and order revealing encryption schemes," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 217–222, 2021.

[48] M. Ristic and B. Noack, "Encrypted fast covariance intersection without leaking fusion weights," in *Proceedings of the 2022 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Bedford, UK, Sep. 2022, pp. 1–6.

[49] C. Funk, B. Noack, and U. D. Hanebeck, "Conservative quantization of fast covariance intersection," in *Proceedings of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Virtual Conference, Sep. 2020, pp. 68–74.

[50] ——, "Conservative quantization of covariance matrices with applications to decentralized information fusion," *Sensors*, vol. 21, no. 9, 2021.

[51] R. Forsling, Z. Sjanic, F. Gustafsson, and G. Hendeby, "Consistent distributed track fusion under communication constraints," in *Proceedings of the 22nd IEEE International Conference on Information Fusion (FUSION)*, Ottawa, Canada, Jul. 2019, pp. 1–8.

[52] ——, "Communication efficient decentralized track fusion using selective information extraction," in *Proceedings of the 23rd IEEE International Conference on Information Fusion (FUSION)*, Virtual Conference, Jul. 2020, pp. 1–8.

[53] ——, "Optimal linear fusion of dimension-reduced estimates using eigenvalue optimization," in *Proceedings of the 25th IEEE International Conference on Information Fusion (FUSION)*, Linköping, Sweden, Jul. 2022, pp. 1–8.

[54] R. Forsling, F. Gustafsson, Z. Sjanic, and G. Hendeby, "Decentralized state estimation in a dimension-reduced linear regression," Oct. 2022, preprint, arXiv. [Online]. Available: https://arxiv.org/abs/2210.06947

[55] J. Kiefer, "Sequential minimax search for a maximum," *Proceedings of the American Mathematical Society*, vol. 4, no. 3, pp. 502–506, 1953.

[56] M. Reinhardt, B. Noack, and U. D. Hanebeck, "Closed-form optimization of covariance intersection for low-dimensional matrices," in *Proceedings of the 15th IEEE International Conference on Information Fusion (FUSION)*, Singapore, Jul. 2012.

[57] J. Wishart, "The generalised product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20A, no. 1/2, pp. 32–52, 1928.

[58] Z. Deng, P. Zhang, W. Qi, J. Liu, and Y. Gao, "Sequential covariance intersection fusion Kalman filter," *Information Sciences*, vol. 189, pp. 293–309, Apr. 2012.

[59] C.-Y. Chong and S. Mori, "Convex combination and covariance intersection algorithms in distributed fusion," in *Proceedings of the 4th IEEE International Conference on Information Fusion (FUSION)*, Montréal, Canada,

Aug. 2001.

[60] 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)*, Edinburgh, Scotland, Jul. 2010.

[61] M. Reinhardt, B. Noack, M. Baum, and U. D. Hanebeck, "Analysis of set-theoretic and stochastic models for fusion under unknown correlations," in *Proceedings of the 14th IEEE International Conference on Information Fusion (FUSION)*, Chicago, IL, USA, Jul. 2011.

[62] J. Ajgl, M. Šimandl, M. Reinhardt, B. Noack, and U. D. Hanebeck, "Covariance intersection in state estimation of dynamical systems," in *Proceedings of the 17th IEEE International Conference on Information Fusion (FUSION)*, Salamanca, Spain, Jul. 2014.

[63] J. Ajgl and O. Straka, "Fusion of multiple estimates by covariance intersection: Why and how it is suboptimal," *International Journal of Applied Mathematics and Computer Science*, vol. 28, no. 3, pp. 521–530, Sep. 2018.

[64] ——, "Covariance intersection in track-to-track fusion without memory," in *Proceedings of the 19th IEEE International Conference on Information Fusion (FUSION)*, Heidelberg, Germany, Jul. 2016.

[65] ——, "Covariance intersection in track-to-track fusion with memory," in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Baden-Baden, Germany, Sep. 2016.

[66] W. Qi, P. Zhang, and Z. Deng, "Robust weighted fusion Kalman filters for multisensor time-varying systems with uncertain noise variances," *Signal Processing*, vol. 99, pp. 185–200, Jun. 2014.

[67] Z. Deng, P. Zhang, W. Qi, G. Yuan, J. Liu, Z. Deng, P. Zhang, W. Qi, Yuan, Gao, and J. Liu, "The accuracy comparison of multisensor covariance intersection fuser and three weighting fusers," *Information Fusion*, vol. 14, no. 2, pp. 177–185, Apr. 2013.

[68] R. Forsling, A. Hansson, F. Gustafsson, Z. Sjanic, J. Löfberg, and G. Hendeby, "Conservative linear unbiased estimation under partially known covariances," *IEEE Transactions on Signal Processing*, vol. 70, pp. 3123–3135, Jun. 2022.

[69] N. A. Carlson, "Federated filter for fault-tolerant integrated navigation systems," in *Proceedings of the 1988 IEEE Position Location and Navigation Symposium (PLANS)*, Orlando, FL, USA, Dec. 1988, pp. 110–119.

[70] ——, "Federated square root filter for decentralized parallel processors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, pp. 517–525, May 1990.

[71] H. Li, F. Nashashibi, and M. Yang, "Split covariance intersection filter: Theory and its application to vehicle localization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1860–1871, Dec. 2013.

[72] H. Li, F. Nashashibi, B. Lefaudeux, and E. Pollard, "Track-to-track fusion using split covariance intersection filter-information matrix filter (SCIF-IMF) for vehicle surrounding environment perception," in *Proceedings of the 16th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Hague, Netherlands, Oct. 2013.

[73] B. Noack, M. Baum, and U. D. Hanebeck, "Automatic exploitation of independencies for covariance bounding in fully decentralized estimation," in *Proceedings of the 18th IFAC World Congress*, Milan, Italy, Aug. 2011.

[74] S. Reece and S. Roberts, "Robust, low-bandwidth, multi-vehicle mapping," in *Proceedings of the 8th IEEE International Conference on Information Fusion (FUSION)*, vol. 2, Philadelphia, PA, USA, Jul. 2005.

[75] 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)*, Baden–Baden, Germany, Aug. 2001, pp. 85–90.

[76] J. Ajgl and O. Straka, "Lower bounds in estimation fusion with partial knowledge of correlations," in *Proceedings of the 2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Karlsruhe, Germany, Sep. 2021, pp. 1–6.

[77] A. Petersen and M.-A. Beyer, "Partitioned covariance intersection," in *Proceedings of 2011 International Symposium Information on Ships*, Hamburg, Germany, Sep. 2011.

[78] J. Ajgl and O. Straka, "Rectification of partitioned covariance intersection," in *Proceedings of the 2019 American Control Conference (ACC)*, Philadelphia, PA, USA, Jul. 2019.

[79] N. R. Ahmed, "Conditionally factorized DDF for general networked Bayesian estimation," in *Proceedings of the 2014 IEEE International Conference on Multisensor Fusion and Information Integration (MFI)*, Beijing, China, Sep. 2014.

[80] N. R. Ahmed, W. W. Whitacre, S. Moon, and E. W. Frew, "Factorized covariance intersection for scalable partial state decentralized data fusion," in *Proceedings of the 19th IEEE International Conference on Information Fusion (FUSION)*, Heidelberg, Germany, Jul. 2016.

[81] N. R. Ahmed, W. Whitacre, S. Moon, and E. W. Frew, "Scalable decentral-

ized partial state estimation with sensor uncertainties using factorized data fusion," in *Proceeding of the AIAA SciTech 2016 Infotech@Aerospace Conference*, San Diego, CA, USA, Jan. 2016.

[82] B. Noack, J. Sijs, and U. D. Hanebeck, "Algebraic analysis of data fusion with ellipsoidal intersection," in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Baden-Baden, Germany, Sep. 2016.

[83] A. R. Benaskeur, "Consistent fusion of correlated data sources," in *Proceedings of the 28th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, vol. 4, Nov. 2002, pp. 2652–2656.

[84] Y. Zhou and J. Li, "Data fusion of unknown correlations using internal ellipsoidal approximation," in *Proceedings of the 17th IFAC World Congress (IFAC 2008)*, Seoul, South Korea, Jul. 2008.

[85] J. Sijs and M. Lazar, "Empirical case-studies of state fusion via ellipsoidal intersection," in *Proceedings of the 14th IEEE International Conference on Information Fusion (FUSION)*, Chicago, IL, USA, Jul. 2011.

[86] ——, "State-fusion with unknown correlation: Ellipsoidal intersection," *Automatica*, vol. 48, no. 8, pp. 1874–1878, Aug. 2012.

[87] J. Nygårds, V. Deleskog, and G. Hendeby, "Safe fusion compared to established distributed fusion methods," in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Baden-Baden, Germany, Sep. 2016.

[88] ——, "Decentralized tracking in sensor networks with varying coverage," in *Proceedings of the 21st IEEE International Conference on Information Fusion (FUSION)*, Cambridge, UK, Jul. 2018.

[89] T.-T. Wu, J. An, C.-S. Ding, and S.-X. Luo, "Research on ellipsoidal intersection fusion method with unknown correlation," in *Proceedings of the 15th IEEE International Conference on Information Fusion (FUSION)*, Singapore, Jul. 2012.

[90] B. Noack, J. Sijs, M. Reinhardt, and U. D. Hanebeck, "Decentralized data fusion with inverse covariance intersection," *Automatica*, vol. 79, pp. 35–41, May 2017.

[91] B. Noack, J. Sijs, and U. D. Hanebeck, "Inverse covariance intersection: New insights and properties," in *Proceedings of the 20th IEEE International Conference on Information Fusion (FUSION)*, Xi'an, China, Jul. 2017.

[92] J. Ajgl and O. Straka, "Analysis of partial knowledge of correlations in an estimation fusion problem," in *Proceedings of the 21st IEEE International Conference on Information Fusion (FUSION)*, Cambridge, UK, Jul. 2018.

[93] ——, “Decentralised estimation with correlation limited by optimal processing of independent data,” *Automatica*, vol. 99, pp. 132–137, Jan. 2019.

[94] ——, “Inverse covariance intersection fusion of multiple estimates,” in *Proceedings of the 23rd IEEE International Conference on Information Fusion (FUSION)*, Virtual Conference, Jul. 2020, pp. 1–8.

[95] T. Sun and M. Xin, “Distributed estimation with iterative inverse covariance intersection,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022.

[96] H. Zhu, Q. Zhai, M. Yu, and C. Han, “Estimation fusion algorithms in the presence of partially known cross-correlation of local estimation errors,” *Information Fusion*, vol. 18, pp. 187–196, Jul. 2014.

[97] S. Leonardos and K. Daniilidis, “A game-theoretic approach to robust fusion and Kalman filtering under unknown correlations,” in *Proceedings of the 2017 American Control Conference (ACC)*, Seattle, WA, USA, Jul. 2017, pp. 2568–2573.

[98] M. Zarei-Jalalabadi, S. M. B. Malaek, and S. S. Kia, “A track-to-track fusion method for tracks with unknown correlations,” *IEEE Control Systems Letters*, vol. 2, no. 2, pp. 189–194, Apr. 2018.

[99] E. Blasch, T. Pham, C.-Y. Chong, W. Koch, H. Leung, D. Braines, and T. Abdelzaher, “Machine learning/artificial intelligence for sensor data fusion–opportunities and challenges,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 7, pp. 80–93, Jul. 2021.

[100] S. Radtke and U. D. Hanebeck, “Learning and exploiting partial knowledge in distributed estimation,” in *Proceedings of the 2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2021)*, Karlsruhe, Germany, Sep. 2021.

[101] C. Chen and S. S. Kia, “Cooperative localization using learning-based constrained optimization,” *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7052–7058, Jul. 2022.

[102] O. Dagan and N. R. Ahmed, “Factor graphs for heterogeneous Bayesian decentralized data fusion,” in *Proceedings of the 24th International Conference on Information Fusion (FUSION)*, Sun City, South Africa, Nov. 2021.

[103] S. Radtke, B. Noack, and U. D. Hanebeck, “Reconstruction of cross-correlations between heterogeneous trackers using deterministic samples,” in *Proceedings of the 1st Virtual IFAC World Congress (IFAC-V 2020)*, July 2020.

[104] X.-R. Li and Z. Zhao, “Measuring estimator’s credibility: Noncredibility index,” in *Proceedings of the 9th IEEE International Conference on Information Fusion (FUSION)*, Florence, Italy, Jul. 2006.

[105] X.-R. Li, Z. Zhao, and V. P. Jilkov,

"Practical measures and test for credibility of an estimator," in *Proceedings of Workshop on Estimation Tracking and Fusion — A Tribute to Yaakov Bar-Shalom*, Monterey, CA, USA, May 2001, pp. 481–495.

Summary

Decentralized sensor networks offer a robust and scalable infrastructure for solving data fusion problems. If done correctly, decentralized data fusion achieves enhanced estimation quality. However, doing it incorrectly leads to unreliable estimates, leaving one with a false sense of estimation accuracy. A common reason for such overconfident estimates is the improper treatment of correlations among the estimates to be fused. Bookkeeping of correlations is cumbersome and often impossible, sparking the invention of covariance intersection a quarter-century ago. In this paper, we take a comprehensive tutorial-style path to revisit covariance intersection by highlighting the challenges in decentralized data fusion, scrutinizing the conservativeness and generality of covariance intersection, following its progress in the past quarter-century, and pointing out recent and future directions.

Sidebar: A Brief Survey of Distributed and Decentralized Estimation

Estimation in distributed and decentralized sensor networks has evolved into a broad field, and various valuable methods have been developed. The main problem is handling cross-correlations between estimates, which leads to the different underlying assumptions for each method. Pioneering work focused on distributed implementations of the Kalman filter. The technical report by [S1] outlines several formulations of the Kalman filter algorithm for managing multisensor data in a distributed setting. The method in [S2] exploits the structure of systems split into interconnected linear dynamical subsystems in order to decompose the Kalman filter, and [S3] has derived an optimal distributed version of the Kalman filter, which is extended in [S4] and [S5] and is similar to the later derived distributed Kalman filter in [S6]. Such methods represent algebraic reformulations of the Kalman filter and, as such, produce optimal estimates. Although some relaxations like [S7] exist, they require specific prerequisites that can limit their use in complex network architectures and, in particular, prohibit a fully decentralized processing in which agents do not know the other agents, their models, and their number beforehand.

A widely used class of DDF algorithms relate to hierarchical systems that implement tracklet or channel-filter fusion [S8]–[S10]. These filters keep track of and explicitly subtract the common information from the fusion result. By storing the state over multiple time steps, correlations induced by process noise can also be addressed to a certain degree [S11], [S12]. Specific solutions for common process noise reconstruct the joint covariances [S13], for which sample-based techniques prove effective [S14], [S15]. Another important class of DDF algorithms is consensus filtering [S16]. Initial consensus schemes did not reliably compute the covariance matrix. By combining consensus filtering with CI, [S17] have derived robust and conservative methods, which have been further developed, for instance, in [S18], to increase robustness to network failures. Similarly, [S19] have integrated CI into diffusion-based distributed Kalman filters to provide conservative estimates.

This brief excursion demonstrates the broad scope of the field and the variety of contributions. For further reading about early contributions in the field of distributed and decentralized filtering, the survey provided in [S20] is suggested. More recent overviews, with a particular focus on target tracking, are found in [S21]–[S24]. Multisensor fusion for robotic systems is discussed in [S25]. A broad review of multisensor data fusion techniques is provided by [S26], which, for instance, includes the handling of imprecise information, outliers, and conflicting data. A review with particular focus on networked systems can be found in [S27]. In particular, research efforts dedicated to network-induced effects like packet delays and losses or quantization are studied. Distributed estimation is a key tool for multisensor data fusion: [S28] presents a broad overview of the underlying theory and the required Bayesian inference techniques. A concise discussion and review of the past forty years of distributed estimation are presented in the work of [S29]. Recent reviews of data fusion under unknown correlations and data inconsistency can be found in [S30], [S31].

References

[S1] D. Willner, C.-B. Chang, and K.-P. Dunn, “Kalman filter configurations for multiple radar systems,” Lexington, MA, USA, pp. 1–61, Apr. 1976, Technical Note.

[S2] M. Hassan, G. Salut, M. Singh, and A. Titli, “A decentralized computational algorithm for the global Kalman filter,” *IEEE Transactions on Automatic Control*, vol. 23, no. 2, pp. 262–268, 1978.

[S3] 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, Apr. 1979.

[S4] A. Willsky, M. Bello, D. Castanon, B. Levy, and G. Verghese, “Combining and updating of local estimates and regional maps along sets of one-dimensional tracks,” *IEEE Transactions on Automatic Control*, vol. 27, no. 4, pp. 799–813, 1982.

[S5] G. J. Bierman and M. R. Belzer, “A decentralized square root information filter/smoother,” in *Proceedings of the 24th IEEE Conference on Decision and Control (CDC)*, Fort Lauderdale, FL, USA, Dec. 1985, pp. 1902–1905.

[S6] F. Govaers and W. Koch, ”Distributed Kalman filter fusion at arbitrary instants of time,” in *Proceedings of the 13th IEEE International Conference on Information Fusion (FUSION)*, Edinburgh, Scotland, Jul. 2010.

[S7] M. Reinhardt, B. Noack, and U. D. Hanebeck, ”The hypothesizing distributed Kalman filter,” in *Proceedings of the 2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pp. 305–312, Hamburg, Germany, Sep. 2012.

[S8] C.-Y. Chong, ”Hierarchical estimation,” in *Proceedings of 2nd MIT/ONR Workshop*, Monterey, CA, USA, Jul. 1979.

[S9] H. R. Hashemipour, S. Roy, and A. J. Laub, “Decentralized structures for parallel Kalman filtering,” *IEEE Transactions on Automatic Control*, vol. 33, no. 1, pp. 88–94, 1988.

[S10] S. Grime and H. F. Durrant-Whyte, ”Data fusion in decentralized sensor networks,” *Control Engineering Practice*, vol. 2, pp. 849–863, Oct. 1994

[S11] S. P. McLaughlin, R. J. Evans, and V. Krishnamurthy, ”Data incest removal in a survivable estimation fusion architecture,” in *Proceedings of the 6th IEEE International Conference on Information Fusion (FUSION)*, Cairns, QL, Australia, Jul. 2003.

[S12] S. P. McLaughlin, V. Krishnamurthy, and S. Challa, ”Managing data incest in a distributed sensor network,” in

Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, China, Apr. 2003.

[S13] M. Reinhardt, B. Noack, and U. D. Hanebeck, "Reconstruction of joint covariance matrices in networked linear systems," in *Proceedings of the 48th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, Mar. 2014.

[S14] J. Steinbring, B. Noack, M. Reinhardt, and U. D. Hanebeck, "Optimal sample-based fusion for distributed state estimation," in *Proceedings of the 19th IEEE International Conference on Information Fusion (FUSION)*, Heidelberg, Germany, Jul. 2016.

[S15] S. Radtke, B. Noack, U. D. Hanebeck, and O. Straka, "Reconstruction of cross-correlations with constant number of deterministic samples," in *Proceedings of the 21st IEEE International Conference on Information Fusion (FUSION)*, Cambridge, UK, Jul. 2015.

[S16] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proceedings of the 44th IEEE Conference on Decision and Control*, Seville, Spain, Dec. 2005, pp. 6698–6703.

[S17] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, and A. Graziano, "Consensus-based linear and nonlinear filtering," *IEEE Transactions on Automatic Control*, vol. 60, no. 5, pp. 1410–1415, May 2015.

[S18] A. Tamjidi, R. Oftadeh, M. N. G. Mohamed, D. Yu, S. Chakravorty, and D. Shell, "Unifying consensus and covariance intersection for efficient distributed state estimation over unreliable networks," *IEEE Transactions on Robotics*, vol. 37, no. 5, Oct. 2021.

[S19] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 891–902, 2012.

[S20] T. H. Kerr and L. Chin, *Advances in the Techniques and Technology of the Application of Nonlinear Filters and Kalman Filters*. Neuilly sur Seine, France: North Atlantic Treaty Organization: The Advisory Group for Aerospace Research and Development, 1982, no. 256, ch. The theory and techniques of discrete-time decentralized filters, pp. 55–93.

[S21] D. Smith and S. Singh, "Approaches to multisensor data fusion in target tracking: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, pp. 1696–1710, Oct. 2006.

[S22] M. E. Liggins II, D. L. Hall, and J. Llinas, eds., *Handbook of Multisensor Data Fusion: Theory and Practice*. The Electronical Engineering and Applied Signal Processing Series, CRC Press, 2 ed., 2009.

[S23] D. L. Hall, C.-Y. Chong, J. Llinas, and M. E. Liggins II, eds., *Distributed Data Fusion for Network-Centric Operations*. The Electronical Engineering and

Applied Signal Processing Series, CRC Press, 2 ed., 2013.

[S24] F. Castanedo, "A review of data fusion techniques," *The Scientific World Journal*, vol. 2013, pp. 1–9, Sep. 2013.

[S25] R. C. Luo, C. C. Chang, and C. C. Lai, "Multisensor fusion and integration: Theories, applications, and its perspectives," *IEEE Sensors Journal*, vol. 11, pp. 3122-3138, Aug. 2011.

[S26] B. Khalegi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, pp. 28–44, Jan. 2013.

[S27] S. Sun, H. Lin, J. Ma, and X. R. Li, "Multi-sensor distributed fusion estimation with applications in networked systems: A review paper," *Information Fusion*, vol. 38, pp. 122–134, Nov. 2017.

[S28] M. E. Campbell and N. R. Ahmed, "Distributed data fusion: Neighbors, rumors, and the art of collective knowledge," *IEEE Control Systems Magazine*, vol. 36, pp. 83–109, Aug. 2016.

[S29] C.-Y. Chong, K. Chang, and S. Mori, "A review of forty years of distributed estimationn," in *Proceedings of the 21st IEEE International Conference on Information Fusion (FUSION)*, Cambridge, UK, Jul. 2018.

[S30] M. A. Bakr and S. Lee, "Distributed multisensor data fusion under unknown correlation and data inconsistency," *Sensors*, vol. 17, pp. 1–31, Oct. 2017

[S31] W. Li and F. Yang, “Information fusion over network dynamics with unknown correlations: An overview,” *International Journal of Network Dynamics and Intelligence*, Feb. 2023.

Sidebar: Practitioner's Point-Of-View

Recall the definition of a conservative estimate

$$\mathbf{P} \succeq \mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\mathsf{T}].$$

It is known that if $\hat{\mathbf{x}}$ and $\mathbf{P}$ are computed using CI then this criterion holds as long as the merged estimates are conservative and unbiased themselves. Simulations and real-world experiments however indicate that CI is often overly conservative [51], [52], [88], [90]. The reason for this is that the extreme cross-correlations assumed in CI seldom occur in typical data fusion problems [90]. Hence, a practitioner that is about to deploy a fusion rule for a given DDF problem should investigate the applicability of different fusion rules and analyze their performance with respect to conservativeness. The scope of this section is to exemplify a simple methodology for how to reason when deciding for a fusion rule. The methodology involves the following steps:

1) Specify sensors, local filters and communication parameters.
2) Specify considered fusion methods and define a metric for conservativeness.
3) Define characteristic target trajectories where $\mathbf{x}_k$ is the state of the target at time k.
4) Tune the local filters for the characteristic trajectories.
5) Using Monte Carlo (MC) simulations, evaluate each fusion method with respect to both performance and conservativeness.

The average normalized estimation error squared (ANEES, [104]) is used to quantify conservativeness. Let $\hat{\mathbf{x}}^i_{k|k}$ be an estimate of $\mathbf{x}_k$ computed by the ith MC run at time k. Let $\mathbf{P}^i_{k|k}$ be the reported covariance of $\hat{\mathbf{x}}^i_{k|k}$. The ANEES ε_k given at time k is defined as

$$\begin{aligned}\varepsilon_k &= \frac{1}{nM}\sum_{i=1}^{M}(\hat{\mathbf{x}}^i_{k|k}-\mathbf{x}_k)^\mathsf{T}(\mathbf{P}^i_{k|k})^{-1}(\hat{\mathbf{x}}^i_{k|k}-\mathbf{x}_k)\\ &= \frac{1}{nM}\sum_{i=1}^{M}(\tilde{\mathbf{x}}^i_{k|k})^\mathsf{T}(\mathbf{P}^i_{k|k})^{-1}\tilde{\mathbf{x}}^i_{k|k}, \qquad \text{(S1)}\end{aligned}$$

where M is the number of MC runs. Ideally, $\varepsilon_k = 1$ but in practice this is seldom the case, especially in DDF problems. Still it is desirable to have $\varepsilon_k \approx 1$. Roughly speaking, a too optimistic covariance matrix is indicated by $\varepsilon_k > 1$ and $\varepsilon_k < 1$ characterizes an overly conservative covariance matrix. If $\mathbb{E}[\tilde{\mathbf{x}}^i_{k|k}] = \mathbf{0}$, then ε_k is approximately chi-square distributed according to $\varepsilon_k \sim \chi^2_{nM}\left(\frac{1}{nM}\right)$ [105], where in this case $\chi^2_{nM}\left(\frac{1}{nM}\right)$ is the central chi-square distribution with nM degrees of freedom scaled by a factor $\frac{1}{nM}$. Since nM becomes large when running many MC simulations $\chi^2_{nM}\left(\frac{1}{nM}\right)$ can fairly accurately be approximated by a Gaussian distribution. For this approximation two-sided confidence intervals can be derived for ε_k. Approximate 95% and 99% confidence intervals are respectively given by [105]

$$\left(\left[1-a-1.96\sqrt{a}\right]^3, \left[1-a+1.96\sqrt{a}\right]^3\right), \qquad \text{(S2a)}$$

$$\left(\left[1-a-2.576\sqrt{a}\right]^3, \left[1-a+2.576\sqrt{a}\right]^3\right), \qquad \text{(S2b)}$$

where $a = \frac{2}{9nM}$. Performance is evaluated using root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\tilde{\mathbf{x}}^i_{k|k})^\mathsf{T}\tilde{\mathbf{x}}^i_{k|k}}.$$

The methodology is exemplified using the same DDF scenario as in the previously used vehicle tracking example. A target vehicle is tracked by two other vehicles. In each vehicle local measurements are filtered into a local estimate of the target using an EKF with a constant velocity model for the dynamics. The local estimates are exchanged and the receiving vehicle merges the received estimate with its own local estimate. The process noise covariance is tuned such that a local EKF (LKF) have ANEES close to 1. The following fusion methods are evaluated: naïve fusion, CI, ICI and the LE method.

The results from $M = 1\,000$ MC simulations are plotted in Figure S1. Only RMSE and ANEES corresponding to the position components have been computed. The RMSE curves in presence of fusion for the most parts are aligned. However, the naïve fusion rule gives a relatively large increase in RMSE at the turning phase, but more importantly severely underestimates the covariance at this phase as indicated by the ANEES plot. The high ANEES is mainly due to double counting of information. CI is on the other side of the spectra as it computes an overly conservative estimate leading to low ANEES. Somewhere in between naïve fusion and CI with respect to ANEES falls ICI and LE. LE in general follows the LKF curve and mainly lies within the 99% confidence interval while at certain cases lies above this interval. ANEES for ICI is typically significantly below 1. Depending on the situation and the specific robustness requirements, ICI or LE would be suggested for this configuration.

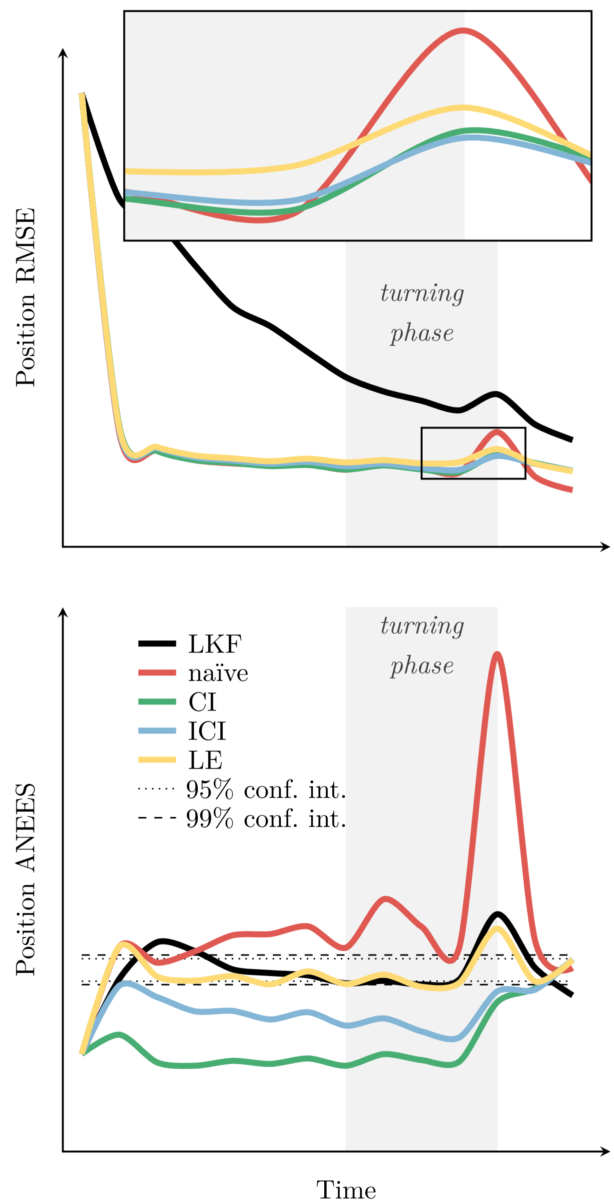


Figure S1: Monte Carlo results of the simulated vehicle tracking scenario. Only quantities related to the position components have been computed. Root mean square error (RMSE) is shown in the top plot. Average normalized estimation error squared (ANEES), as defined in (S1), is shown in the bottom plot where also 95% and 99% confidence intervals are included.

It should be emphasized that the outcome of this type of simulation is strongly influenced on the communication pattern and rate. If estimates are exchanged very seldom then it could be the case that even a naïve fusion method would yield ANEES values close to 1. As always it is important to understand the problem at hand when choosing a method.

Sidebar: Conservative Non-Gaussian Fusion

The Kalman filter does not require the noise to be Gaussian to be a linear MSE optimal estimator. The Gaussianity assumption, however, comes into play to derive the Kalman filter from Bayes' rule [S32] with all estimates and covariances interpreted as parameters of Gaussian probability densities and models affected by Gaussian errors. CI possesses a corresponding relationship: On the one hand, the conservativeness of CI holds in linear fusion problems and relates to the MSE without any assumption on Gaussianity. On the other hand, an alternative interpretation of CI is the *exponential mixture density* (EMD)

$$p_{\mathrm{CI}}(\mathbf{x}) = \frac{p_1(\mathbf{x})^{\omega} \cdot p_2(\mathbf{x})^{(1-\omega)}}{\int p_1(\mathbf{x})^{\omega} \cdot p_2(\mathbf{x})^{(1-\omega)}\, d\,\mathbf{x}}, \quad \text{(S3)}$$

where p_1 is a Gaussian density with mean $\hat{\mathbf{x}}^1$ and covariance $\mathbf{P}^1$, and p_2 has the corresponding parameters $\hat{\mathbf{x}}^2$ and $\mathbf{P}^2$. The resulting parameters for $p_{\mathrm{CI}}(\mathbf{x})$ then equal the CI estimate (5). This density representation in (S3) has been independently discovered in [S33] and [S34]. The final step from the CI's Gaussian representation toward conservative nonlinear fusion is simple: One can apply (S3) to estimates with arbitrary non-Gaussian representations $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$. A 1D example of this is provided in Figure S2, where p_{CI} for three values of ω has been computed.

The EMD fusion rule has experienced much attention in the literature, which has also established other terms for (S3) such as *normalised weighted geometric mean* [S35],

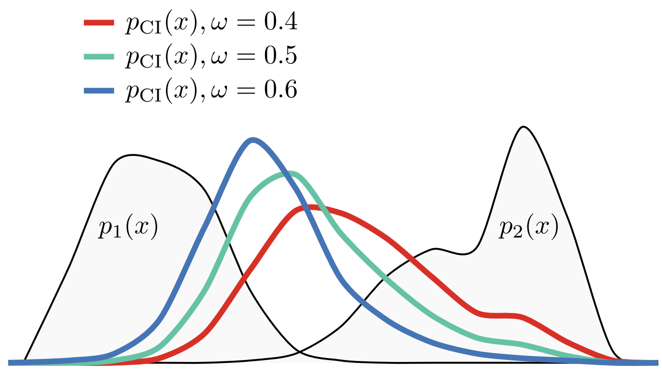


Figure S2: A 1D example of nonlinear fusion. Two arbitrary densities $p_1(x)$ and $p_2(x)$ are fused using the exponential mixture density given in (S3). The density $p_{\mathrm{CI}}(x)$ is computed for three values of ω. Roughly speaking, the parameter ω decides the extent to which each of $p_1(x)$ and $p_2(x)$ are included in $p_{\mathrm{CI}}(x)$.

Chernoff fusion [S36], [S37], and *log-linear pooling* [S38]. Its generalization to multiple densities is discussed in [S39] leading to the weighted Kullback-Leibler average [S40] of the densities $p_1(\mathbf{x}), \ldots, p_N(\mathbf{x})$ given by

$$p(\mathbf{x}) = \frac{\prod_{i=1}^{N} p_i(\mathbf{x})^{\omega_i}}{\int \prod_{i=1}^{N} p_i(\mathbf{x})^{\omega_i} d\mathbf{x}},$$

with $\omega_i \geq 0$ and $\sum_{i=1}^{N} \omega_i = 1$. In the Gaussian case, CI can hence be interpreted as the Kullback-Leibler average. Moreover, this result is important when it comes to consensus on general probability densities.

Other conservative methods from section *Derivatives and Alternatives to Covariance Intersection* correspondingly have led to non-Gaussian generalizations. A nonlinear variant of the federated Kalman filter has been developed in [S41]. Analogous to split CI, EMD can handle dependent and independent parts separately. Common information shared by two probability densities can be removed with the nonlinear channel filter [S42] and log-density formulations [S43]. A nonlinear generalization of ICI has been derived in [S44].

While the Kalman filter and CI are applicable for fusion of non-Gaussian estimates, the resulting estimates are only parameterized by mean and covariance. In applications where such an incomplete uncertainty is unacceptable, the formula in (S3) should be used instead of the original formulation of CI in (5). As a consequence, non-Gaussian generalizations of CI are building up momentum, which is particularly fueled by the demand in multisensor multiobject tracking applications, see, for instance, [S45]–[S47]. However, although (S3) looks more general than the CI formulas (5), the universality is deceptive: The user must not underestimate the computational complexity of applying (S3) to non-Gaussian densities and, most importantly, a common misconception about EMD fusion is that it inherits CI's notion of conservativeness.

The fusion of Gaussian mixture densities demonstrates the computational challenges of applying EMD fusion (S3), which includes exponentiation, multiplication and division [S48]. In [S49] and [S50] approximations for the powers of Gaussian mixture densities are proposed. A different approach is to apply CI compenentwise [S51], which has been further studied and generalized in [S52]. State-space transformations can also simplify the fusion process in specific scenarios [S53]. Particle representations also render EMD fusion challenging as two particle sets cannot simply be multiplied. Typical solutions reapproximate either p_1 or p_2 by a Gaussian or Gaussian mixture density [S54], [S55]. Another challenge is the computation of an appropriate weighting parameter, which is also more challenging than for CI. The selection of ω in (S3) typically relies on information-theoretic measures [S56], [S57].

Nonlinear estimation is undisputably in need of conservative fusion methods, but the question remains: To which extent does the EMD in (S3) guarantee conservativeness? As noticed in [S58], [S59], EMD fusion inherits from CI that double-counting is avoided, and common information is only incorporated once. Beyond this property, a notion of conservativeness like (1) for CI is missing. An initial approach to define conservative densities has been proposed by [S35], which uses differential entropy and a notion of order preservation. The use of entropy is further studied in [S60]. Minimum-volume sets have recently been suggested by [S61] as an effective notion of conservativeness, which circumvents shortcomings of the entropy-based notion and has a strong intuitive backing. Another recent work [S38] reviews the fusion of probability densities. These works take big steps toward answering the question posed at the beginning of this paragraph. Yet, it remains to be studied how to characterize and parameterize dependencies among p_1 and p_2 and how p_{CI} can account for them. The fusion of non-Gaussian estimates lacks a clear definition of unknown dependencies, while CI addresses the well-defined set of unknown cross-correlations.

References

[S32] J. Uhlmann and S. J. Julier, ”Gaussianity and the Kalman filter: A simple yet complicated relationship,” *Journal de Ciencia e Ingeniería*, vol. 14, pp. 21–26, Jun. 2022.

[S33] M. B Hurley, ”An information theoretic justification for covariance intersection and its generalization,” in *Proceedings of the 5th IEEE International Conference on Information Fusion (FUSION)*, Annapolis, MD, USA, Jul. 2002.

[S34] R. P. S. Mahler, ”Optimal/robust distributed data fusion: A unified approach,” in *Proceedings of SPIE*, vol. 4052, 2000.

[S35] T. Bailey, S. J. Julier, and G. Agamemnnoni, ”On conservative fusion of information with non-Gaussian dependence,” in *Proceedings of the 15th IEEE International Conference on Information Fusion (FUSION)*, Singapore, Jul. 2012.

[S36] W. J. Farrell and C. Ganesh, ”Generalized Chernoff fusion approximation for practical distributed data fusion,” in *Proceedings of the 12th IEEE International Conference on Information Fusion (FUSION)*, Seattle, WA, USA, Jul. 2009.

[S37] M. Günay, U. Orguner, and M. Demirekler, ”Chernoff fusion of Gaussian mixtures based on sigma-point approximation,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, pp. 2732–2746, Dec. 2016

[S38] G. Koliander, Y. El-Laham, P. M. Djurić, and F. Halawatsch, ”Fusion of probability density functions,” *Proceedings of the IEEE*, vol. 110, pp. 404–453, Apr. 2022

[S39] G. Battistelli and L. Chisci, “Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability,” *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.

[S40] A. E. Abbas, “A Kullback-Leibler view of linear and log-linear pools,” *Decision Analysis*, vol. 6, no. 1, pp. 25–37, Mar. 2009.

[S41] B. Noack, S. J. Julier, M. Reinhardt, and U. D. Hanebeck, ”Nonlinear federated filtering,” in *Proceedings of the 16th IEEE International Conference on Information Fusion (FUSION)*, Istanbul, Turkey, Jul. 2013.

[S42] F. Bourgault and H. F. Durrant-Whyte, ”Communication in general decentralized filters and the coordinated search strategy,” in *Proceedings of the 7th IEEE International Conference on Information Fusion (FUSION)*, Stockholm, Sweden, Jun. 2004.

[S43] B. Noack, D. Lyons, M. Nagel and U. D. Hanebeck, ”Nonlinear information filtering for distributed multisensor data fusion,” in *Proceedings of the 2011 American Control Conference (ACC)*, San Fransisco, CA, USA, Jun. 2011.

[S44] B. Noack, U. Orguner, and U. D. Hanebeck, ”Nonlinear decentralized data fusion with generalized inverse covariance intersection," in *Proceedings of the 22nd IEEE International Conference on Information Fusion (FUSION)*, Ottawa, Canada, Jul. 2019.

[S45] D. E. Clark, S. J. Julier, R. P. S. Mahler, and B. Ristic, "Robust multi-object sensor fusion with unknown correlations," in *Proceedings of the 2010 IEEE International Conference In Sensor Signal Processing for Defence (SSPD)*, London, UK, Sep. 2010

[S46] M. Üney, D. E. Clark, and S. J. Julier, "Distributed fusion of PHD filters via exponential mixture densities," *IEEE Selected Topics in Signal Processing*, vol. 7, pp. 521–531, Jun. 2013.

[S47] C. Fantacci, B.-N. Vo, B.-T. Vo, G. Battistelli, and L. Chisci, "Robust fusion for multisensor multiobject tracking," *IEEE Signal Processing Letters*, vol. 25, pp. 640–644, May. 2018.

[S48] N. R. Ahmed, "What's one mixture divided by another?: A unified approach to high-fidelity distributed data fusion with mixture models," in *Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, San Diego, CA, USA, Sep. 2015.

[S49] J. Ajgl, M. Šimandl, and J. Duník, "Approximation of powers of Gaussian mixtures," in *Proceedings of the 18th IEEE International Conference on Information Fusion (FUSION)*, Washington D.C., USA, Jul. 2015.

[S50] H. Zhu, K. Guo, and S. Chen, "Fusion of Gaussian mixture models for maneuvering target tracking in the presence of unknown cross-correlation," *Chinese Journal of Electronics*, vol. 25, pp. 270–276, Mar. 2016.

[S51] 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 IEEE International Conference on Information Fusion (FUSION)*, Florence, Italy, Jul. 2006.

[S52] B. Noack, M. Reinhardt, and U. D. Hanebeck, "On nonlinear track-to-track fusion with Gaussian mixtures," in *Proceedings of the 17th IEEE International Conference on Information Fusion (FUSION)*, Salamanca, Spain, Jul. 2014.

[S53] B. Noack, M. Baum, and U. D. Hanebeck, "Covariance intersection in nonlinear estimation based on pseudo Gaussian densities," in *Proceedings of the 14th IEEE International Conference on Information Fusion (FUSION)*, Chicago, IL, USA, Jul. 2011.

[S54] O. Hlinka, O. Slučiak, F. Hlawatsch, and M. Rupp, "Distributed data fusion using iterative covariance intersection," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014.

[S55] O. Tslil and A. Carmi, "Information fusion using particles intersection," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, Apr. 2018.

[S56] F. Nielsen, "An information-geometric characterization of Chernoff information," *IEEE Signal Processing Letters*, vol. 20, pp. 269–272, Mar. 2013.

[S57] D. Clarke, "Minimum information loss fusion in distributed sensors networks," in *Proceedings of the 19th IEEE International Conference on Information Fusion (FUSION)*, Heidelberg, Germany, Jul. 2016.

[S58] S. J. Julier, T. Bailey, and J. K. Uhlmann, "Using exponential mixture models for suboptimal distributed data fusion," in *Proceedings of the 2006 IEEE Nonlinear Statistical Signal Processing Workshop (NSSPW)*, Cambridge, UK, pp. 160–163, Sep. 2006.

[S59] S. J. Julier, "Fusion without independence (keynote abstract)," in *Proceedings of the IET Seminar on Tracking and Data Fusion: Algorithms and Applications*, Birmingham, UK, pp. 1–5, Apr. 2008.

[S60] J. Ajgl and M. Šimandl, "Conservativeness of estimates given by probability density functions: Formulation and aspects," *Information Fusion*, vol. 20, pp. 117–128, Nov. 2014.

[S61] S. Lubold and C. N. Taylor, "Formal definitions of conservative probability distribution functions (PDFs)," *Information Fusion*, vol. 88, pp. 175–183, Dec. 2022.