Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Information Fusion 4 (2003) 201–215

INFORMATION FUSION

www.elsevier.com/locate/inffus

Covariance consistency methods for fault-tolerant distributed data fusion

Jeffrey K. Uhlmann *

201 EBW, Department of Computer Engineering and Computer Science, University of Missouri—Columbia, Columbia, MO 65211, USA

Received 2 May 2002; received in revised form 21 August 2002; accepted 22 September 2002

Abstract

This paper presents a general, rigorous, and fault-tolerant framework for maintaining consistent mean and covariance estimates in an arbitrary, dynamic, distributed network of information processing nodes. In particular, a solution is provided that addresses the information deconfliction problem that arises when estimates from two or more different nodes are determined to be inconsistent with each other, e.g., when two high precision (small covariance) estimates place the position of a particular object at very different locations. The challenge is to be able to resolve such inconsistencies without having to access and exploit global information to determine which of the estimates is spurious. The solution proposed in this paper is called Covariance Union.

Keywords: Control; Convex optimization; Covariance Intersection; Covariance Union; Data fusion; Deconfliction; Distributed processing; Fault tolerance; Kalman filter; Network-centric; Network management; Semidefinite programming

1. Introduction

This paper presents a general, rigorous, and fault-tolerant framework for maintaining consistent mean and covariance estimates in an arbitrary, dynamic, distributed network of information processing nodes. In particular, a solution is provided that addresses the information deconfliction problem [18] that arises when estimates from two or more different nodes are determined to be inconsistent with each other, e.g., when two high precision (small covariance) estimates place the position of a particular object at very different locations. The challenge is to be able to resolve such inconsistencies without having to access and exploit global information to determine which of the estimates is spurious. The solution proposed in this paper is called Covariance Union (CU).

The outline of the paper is as follows: Section 2 describes the motivation for developing modular, decentralized data fusion networks and discusses the challenges that they present. Section 3 describes some of the traditional motivations for representing information in the form of mean and covariance estimates. Section 4 describes the Kalman filter and the limitations of its use for distributed data fusion. Section 5 describes the Covariance Intersection (CI) data fusion mechanism and explains how it circumvents the limitations of the Kalman filter. Section 6 discusses the limitations of CI with respect to fault tolerance, and the CU mechanism is introduced to address those limitations. Section 7 provides a detailed example of how the combination of CI and CU supports general, fault-tolerant, distributed data fusion. Section 8 summarizes the contributions of the paper and their applications to related problems.

2. Modularity and distributed data fusion

Modular design is one of the most established and successful methodologies in engineering for dealing with the complexities of extremely large systems. The basic idea is to construct complex systems from well-defined functional components (modules) that can be interconnected through relatively simple interfaces. Some of the benefits of modular design include:

1. The ability to decompose a large design problem into smaller design problems that are separately simpler and more manageable.
2. The ability to localize a system problem by testing and verifying the performance of its component modules in isolation.

* Tel.: +1-573-884-2129.
E-mail address: uhlmannj@missouri.edu (J.K. Uhlmann).

1566-2535/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S1566-2535(03)00036-8

3. The ability to replace an individual module with an improved (though functionally equivalent) module without regard to the global structure of the system of which it is a component.
4. The ability to construct increasingly more complex and powerful modules from simpler modules to support the design of complex systems at higher levels of abstraction.
5. The ability to incorporate redundant functional modules to provide robustness to component failures.
6. The ability to incrementally construct, improve, or adapt a large system by introducing or changing only a small number of components.

Although the above benefits are readily achieved in a wide variety of physical systems (examples range from articles of clothing to hydro-electric dams), they are difficult to attain in the case of modular, or *distributed*, information processing systems [8,29].

In a distributed data processing system, modules perform information processing functions and are connected together by communication links. An example of such a system is a distributed sensing network in which different sensor nodes exchange measured information about an object or process of interest [7,19,29]. In principle each sensor node should be able to incorporate the information from other nodes to construct a better estimate of the state of the object or process than what is possible from its local sensor information. More generally, modularity should be expected to provide:

1. An ability to easily incorporate more sensor modules into the system to produce improved estimates of the state of the observed object or process.
2. An ability of the system to function after the loss of a sensor module (or link) with a performance degradation commensurate with the quality of the information produced by the lost sensor.

Unfortunately, the above characteristics can be difficult to achieve [3,8,29]. The reason for this is that although the information processing nodes may only communicate locally with a small number of other nodes, the information that is processed may propagate globally throughout the system. Consequently, information that is processed at a particular node and communicated to the network may subsequently be received again by that node in some form, so feedback effects determined by the global connectivity of the network must be considered. The problem is that such global inter-dependencies can violate critical conditions necessary to realize the benefits of modularity and decentralization [7,13,19,22,29]. In order to understand these conditions, it is necessary to understand how information is commonly represented, processed, and combined/fused in large-scale applications. These are the subjects of the following two sections.

3. Information, uncertainty, and data fusion

Determining how to represent information and uncertainty is a key first step that impacts all aspects of the data fusion problem. The vast literature on filtering, control, and data fusion includes many Bayesian approaches for fusing information represented in the form of probability distributions or probability density functions (PDFs). The principal difficulty associated with such approaches is that in practice it is rarely if ever possible to use a fixed number of parameters to represent the complete statistical relationship that exists among a set of state variables necessary for the application of Bayes' rule. [1] Consequently, Bayesian methods either incur unbounded computational demands or must apply approximations. Unfortunately, deriving approximations with rigorous performance guarantees seems to be no less challenging, so Bayesian methods are typically applied in violation of assumptions that are critical for mathematical rigor and then supplemented with heuristic tuning mechanisms.

Although heuristic methods can be empirically tailored to function effectively in certain restricted applications, there are many contexts (e.g., novel, stressing, safety-critical) in which rigorous performance guarantees must be maintained. This has motivated attempts to develop alternative formalisms for representing information. One alternative to maintaining a full PDF is to determine specific properties of an unknown error distribution that can be maintained with a fixed number of parameters as new information is incorporated (or fused). Absolute error bounds are statistical properties that can be used to characterize uncertainty without a complete PDF. For example, if the uncertainty associated with each measurement of the (x, y) position of an object has known error bounds in each coordinate, i.e., defines a rectangular uncertainty region, then a fused estimate can be constructed by determining the rectangular intersection of the set of measurements. What is important to note is that the rectangular fused estimate is represented with the same constant number of parameters as the measurements it summarizes, and the

[1] Unless the PDFs associated with n measurements are defined in the same state space and are of a special form, e.g., Gaussian, then the conditional PDF can be expected to require $O(n)$ parameters to specify exactly. It is not generally possible to "compress" the PDF parameterization; however, it is possible to maintain a discretized approximation. This approach was developed in the 1960s [5], but computational costs prevented it from being practically applied until the 1990s [23].

error bounds are guaranteed to be faithful if the bounds associated with each measurement are faithful.

If the uncertainty associated with measured information can be effectively represented in the form of a bounded region in the state space of interest, then a fully rigorous and practical data fusion mechanism is available. The problem is that there is no general empirical method for determining *maximum possible* errors needed to establish the bounds. If the errors are approximately Gaussian distributed, for example, then taking larger and larger sequences of measurements of the error process will lead to larger and larger estimated bounds. What is needed is a property of the unknown distribution that is empirically measurable.

An estimate of the squared error, or *covariance*, associated with measurements from a particular sensor *can* be modeled empirically by examining measurements taken of reference objects whose true states are known. This permits an error covariance matrix to be associated with subsequent measurements of objects whose true states are not known. For example, the measured position of an object in two dimensions can be represented as a vector $\mathbf{a}$ consisting of the object's estimated mean position, e.g., $\mathbf{a} = [\mathbf{x}, \mathbf{y}]^{\mathrm{T}}$, and an error covariance matrix $\mathbf{A}$ that expresses the uncertainty associated with the estimated mean. If the error in the estimated mean vector is denoted as $\tilde{\mathbf{a}}$, then the error covariance matrix is an estimate of the expected squared error, $E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}]$. The estimate is said to be consistent (or conservative) if and only if $\mathbf{A} \geqslant E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}]$ or, equivalently, $\mathbf{A} - E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}]$ is positive definite or semidefinite (i.e., has no negative eigenvalues). [2] The full estimate of a target's state is given by the mean and covariance pair $(\mathbf{a}, \mathbf{A})$.

An error covariance matrix can be used to summarize the contributions of a wide variety of different sources of estimate uncertainty. For example, if the position of a person floating at sea after an accident is reported by the pilot of a search aircraft, a mean and covariance position estimate can be constructed by combining the range and bearing estimates given by the pilot with estimates of the aircraft's absolute position and orientation from onboard systems. Covariance estimates are readily available from most GPS and inertial sensors and can (after appropriate transformations) be summed, but the covariance contribution from the pilot's report may be more difficult to assess. In some applications it may be derived doctrinally from ensemble averages of the historical performance of pilots in training and exercise experiments and then scaled to account for other factors such as inexperience, environmental conditions, and the quality of the observation, e.g., a long and focused observation versus a brief observation under duress. In other applications the covariance may be derived from the pilot's own estimate of positional uncertainty. [3] In the end, however, the only requirement for estimate consistency is that the sum of all error covariances never underestimates the actual squared errors associated with the mean estimate.

The use of covariance matrices to represent uncertainty provides:

1. A measure of uncertainty that can be empirically determined for practical measuring devices/sensors, i.e., it does not have to be assumed a priori as in the case of the bounded error representation.
2. A measure of uncertainty that is extremely general, i.e., is defined for almost any practical error distribution, as opposed to absolute bounds that often do not exist.
3. A measure of uncertainty that has a fixed number of parameters depending only on the dimensionality of the state space of interest.
4. A measure of uncertainty that can be linearly transformed without any other knowledge about an underlying PDF, e.g., if the mean of an estimate $(\mathbf{a}, \mathbf{A})$ is transformed as $\mathbf{Ta}$, then the covariance is $\mathbf{TAT}^{\mathrm{T}}$ [1,17].
5. A measure of uncertainty that can be maintained in a conservative "upper bound" form that avoids the need to determine *exact* statistics about an unknown error distribution [11].

In addition to the above characteristics, the mean and covariance representation of information is also extremely flexible: the mean vector may consist of purely quantitative variables that are measured or estimated, or it may contain descriptive attributes assigned by a human. For example, in a health monitoring application the state variables may relate to the heart rate, blood pressure, and blood oxygen levels of a patient undergoing anesthesia. In a law enforcement application, a criminal suspect database may contain vectors of attributes from eyewitness descriptions that include variables such as height, weight, and colors of eyes, skin, and hair. The entries in the error covariance matrix for a given suspect description not only encode the uncertainty (variance) for each of the variables (height, weight, etc.),

[2] Note that the terms "mean" and "covariance" as defined here do not refer to the same statistical entities assumed in Bayesian derivations of the Kalman filter [17]. In this paper there is no statistical property associated with mean and covariance estimates beyond the definition of consistency. Information on the positive definite ordering can be found in [9].

[3] It is common for pilots to associate *probability containment ellipses* with their position estimates. For example, a 95% containment ellipse is interpreted as the region in which the observed object is expected to be found with a 0.95 probability. Probability containment ellipses can be directly converted into covariance matrices. (An example of the use of containment ellipses and covariances for multi-source error analysis in civilian air applications can be found in [30].)

they also encode information about their relationships. For example, if a suspect is described as having a "medium build", then a relationship is defined between the height and weight variables. Therefore, if a potential suspect is slightly taller than the height estimated by the witness, then the weight should also be slightly greater than the estimate of the witness to be consistent with the description of a medium build. [4] Such covariant relationships are encoded in the cross terms of the covariance matrix.

In some applications uncertainty cannot be adequately represented in the form of a single mean and covariance estimate, e.g., when the true state of the object or process of interest is determined to be in one of multiple discrete or localized "modes". Such cases of multi-modal uncertainty are commonly addressed using *track splitting* or *multiple hypothesis* methods that associate a mean and covariance estimate with each mode. Extremely sophisticated theoretical and practical frameworks have been developed for maintaining and updating mean and covariance estimates in multiple hypothesis tracking systems [1,2,15,20,21,26,28]. Until recently, virtually all such systems depended on a single mechanism for fusing multi-source mean and covariance information: the Kalman filter.

4. The Kalman filter and distributed data fusion

Given two mean and covariance estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$, the data fusion problem of interest in this paper consists of determining a fused estimate $(\mathbf{c}, \mathbf{C})$ that is guaranteed to be consistent and summarizes the information in the two estimates with error (in terms of the size of $\mathbf{C}$) that is less than or equal to that of either estimate. If the two estimates are consistent and presumed to be statistically independent, then a joint estimate can be constructed as

$$\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \right). \tag{1}$$

Letting $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ denote the errors in the respective mean estimates, the key property of the joint covariance estimate is that it satisfies

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \geqslant \begin{bmatrix} E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}] & \mathbf{0} \\ \mathbf{0} & E[\tilde{\mathbf{b}}\tilde{\mathbf{b}}^{\mathrm{T}}] \end{bmatrix}, \tag{2}$$

where the RHS matrix represents the true but unknown joint error covariance, which has zero cross covariance, $E[\tilde{\mathbf{a}}\tilde{\mathbf{b}}^{\mathrm{T}}] = \mathbf{0}$, due to the assumption of statistical independence. The estimated joint covariance is a conservative estimate of the true joint covariance because in practice $\mathbf{A} > E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}]$ and $\mathbf{B} > E[\tilde{\mathbf{b}}\tilde{\mathbf{b}}^{\mathrm{T}}]$. The latter inequalities hold by design in that intentional efforts are made to ensure that estimate error covariances do not underestimate the actual squared errors associated with sensor and kinematic models.

Given a consistent joint covariance for two given d-dimensional estimates, the Kalman filter defines the optimal linear projection of the $2d$-dimensional joint estimate back to the d-dimensional state space of interest. The result of the Kalman projection is a mean and covariance estimate $(\mathbf{c}, \mathbf{C})$ that represents the optimal fusion of the two given mean and covariance estimates. In fact, if there is no additional information available (e.g., distribution information), then the Kalman fusion estimate is optimal according to virtually any error criteria [17].

In the case of statistically independent estimates $(\mathbf{a}_1, \mathbf{A}_1), (\mathbf{a}_2, \mathbf{A}_2), \ldots, (\mathbf{a}_n, \mathbf{A}_n)$, the Kalman fusion equations have a particularly simple form [17]

$$\mathbf{C} = \left(\mathbf{A}_1^{-1} + \mathbf{A}_2^{-1} + \cdots + \mathbf{A}_n^{-1}\right)^{-1}, \tag{3}$$

$$\mathbf{c} = \mathbf{C}\left(\mathbf{A}_1^{-1}\mathbf{a}_1 + \mathbf{A}_2^{-1}\mathbf{a}_2 + \cdots + \mathbf{A}_n^{-1}\mathbf{a}_n\right). \tag{4}$$

If its underlying assumptions hold (i.e., consistency and independence), then the above Kalman equations ensure that the fused estimate $(\mathbf{c}, \mathbf{C})$ is consistent, and $\mathbf{C} \leqslant \mathbf{A}_i \ \forall i, \ 1 \leqslant i \leqslant n$. However, any presumption of statistical independence in practical data fusion contexts should be carefully considered. Specifically, virtually any sensor is subject to time-correlated errors induced by the particular conditions of its use (e.g., changes in temperature, platform vibrations, relative humidity), and errors associated with the non-linear transformation of its measurements (e.g., from local spherical coordinates to a global coordinate frame) are deterministic and therefore non-independent.

If estimates $(\mathbf{a}_1, \mathbf{A}_1), (\mathbf{a}_2, \mathbf{A}_2), \ldots, (\mathbf{a}_n, \mathbf{A}_n)$ are each consistent, but not completely uncorrelated, then it is possible for the Kalman fused estimate $(\mathbf{c}, \mathbf{C})$ to be *inconsistent*. In fact, if $\mathbf{A}_i = \tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^{\mathrm{T}} \ \forall i, \ 1 \leqslant i \leqslant n$, then *any* degree of correlation guarantees inconsistency, i.e., $\mathbf{C} \not\geqslant E[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^{\mathrm{T}}]$. The key point is that a Kalman fused estimate is not guaranteed to be consistent even if each of its given estimates are consistent. The reason why the Kalman filter fails is because although two given estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$ may be individually consistent, the implicit joint covariance may fail to be if independence is assumed when the cross covariance between the estimates is $\mathbf{X} = E[\tilde{\mathbf{a}}\tilde{\mathbf{b}}^{\mathrm{T}}] \neq \mathbf{0}$. Specifically

$$\begin{bmatrix} E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}] & \mathbf{0} \\ \mathbf{0} & E[\tilde{\mathbf{b}}\tilde{\mathbf{b}}^{\mathrm{T}}] \end{bmatrix} \not\geqslant \begin{bmatrix} E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}] & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & E[\tilde{\mathbf{b}}\tilde{\mathbf{b}}^{\mathrm{T}}] \end{bmatrix}. \tag{5}$$

In other words, the Kalman filter fails to produce a consistent fused estimate only when the implicit *joint* estimate is inconsistent. Although the Kalman filter

[4] The Mahalanobis distance measure, discussed in Section 6, can be used to assess similarity between mean and covariance estimates.

equations are very simple and elegant for uncorrelated estimates, more complicated equations can be derived to accommodate any consistent joint covariance with known cross covariance $\mathbf{X} \neq \mathbf{0}$ [11,17]. Therefore, the Kalman filter can be used to produce consistent estimates as long as the given estimates are consistent and their cross covariance is known. Unfortunately, this poses significant challenges. The first challenge is that the cross covariance information must in principle be determined *exactly*. This can be seen from an examination the difference between two joint covariance matrices with different cross terms

$$\begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix} - \begin{bmatrix} \mathbf{A} & \mathbf{Y} \\ \mathbf{Y}^{\mathrm{T}} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{X}-\mathbf{Y} \\ (\mathbf{X}-\mathbf{Y})^{\mathrm{T}} & \mathbf{0} \end{bmatrix}. \tag{6}$$

The difference matrix is not positive semidefinite for any case in which $\mathbf{X} \neq \mathbf{Y}$. The need for absolutely perfect cross covariance information presents difficulties when estimates are the products of non-linear operations (e.g., coordinate transformations, kinematic time projections, human-derived estimates) because the error processes are not perfectly modeled. For example, the same approximate non-linear transformation equations may be applied to convert different radar observations of an object to a common coordinate frame, so the errors committed are clearly not independent, but it may not be possible to determine exact cross covariances.

The other challenge posed by the Kalman filter's need for exact cross covariance information is exemplified by the general distributed data fusion application described in Section 2. Specifically, if nodes in a network are permitted to freely receive, fuse, and transmit estimates, then any estimate received at a node may be correlated to an arbitrary extent with the node's local estimate of the state of the object or process of interest. This situation precludes use of the Kalman filter [13,29].

Many attempts have been made to circumvent the problems associated with the Kalman filter in distributed data fusion contexts by limiting the proliferation of redundant information [19,29]. One approach is to only allow the propagation of raw sensor estimates that are uniquely tagged or otherwise controlled so that nodes can avoid redundantly incorporating the same estimate [19]. Unfortunately, such an approach eliminates the ability to distribute *derived* information. For example, each node may have local capabilities that allow it to extract more information from a set of received sensor measurements than what is possible at other nodes. If composite estimates cannot be transmitted, then no other node will be able to exploit the derived information.

An intuitively appealing approach for permitting the propagation of composite estimates is to associate *lineage* or *pedigree* information with each estimate that allows its cross covariance with other estimates to be determined before fusion with a Kalman filter [22]. However, it has been proven that in a general network (e.g., one with cycles), any such scheme will succumb to combinatorial computational complexity [29]. The conclusion of the analysis in [29] is that no Bayesian-based data fusion mechanism, nor any other mechanism that relies on correlation information, can be used in a general decentralized data fusion network. This conclusion suggests the need for an alternative fusion mechanism that does not require information about the statistical relationships that exist among consistent estimates.

5. Covariance Intersection and distributed data fusion

It turns out that the pessimistic analysis in [29] regarding the application of the Kalman filter for general decentralized data fusion hinges on an implicitly assumed formulation of the Kalman fusion operation. Specifically, it is assumed that when the Kalman filter is to be applied to combine two estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$, the relevant joint covariance must be of the form

$$\begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix}, \tag{7}$$

where the cross covariance $\mathbf{X}$ is either known exactly or else the Kalman filter equations cannot be consistently applied. However, if it is possible to define a joint covariance $\mathbf{M}$ with diagonal blocks $\mathbf{M}_A > \mathbf{A}$ and $\mathbf{M}_B > \mathbf{B}$ such that

$$\mathbf{M} \geqslant \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix} \tag{8}$$

for *any possible* instantiation of the unknown cross covariance $\mathbf{X}$, then the block components of $\mathbf{M}$ can be incorporated into the structure of the Kalman filter equations to yield a provably consistent fused estimate $(\mathbf{c}, \mathbf{C})$. The critical question is whether a provably consistent joint covariance $\mathbf{M}$ can be constructed that yields a useful fused estimate, i.e., one with less associated uncertainty. It turns out that it is possible, and the resulting data fusion mechanism is called *Covariance Intersection* (CI) [25].

CI operates by determining the consistent joint covariance $\mathbf{M}$ for which the Kalman filter equations yield the best possible fused estimate $(\mathbf{c}, \mathbf{C})$ according to a fixed measure of covariance size, e.g., minimum determinant. A particular covariance measure is necessary because there is not generally a unique consistent joint covariance that is minimal in the positive definite ordering. For example, there may exist two consistent joint covariances $\mathbf{M}_1$ and $\mathbf{M}_2$ such that $\mathbf{M}_1 \not\leqslant \mathbf{M}_2$, $\mathbf{M}_2 \not\leqslant \mathbf{M}_1$, and no other consistent joint covariance $\mathbf{M}_3$ exists such that $\mathbf{M}_3 \leqslant \mathbf{M}_1$, $\mathbf{M}_3 \leqslant \mathbf{M}_2$. Therefore, if a fused estimate $(\mathbf{c}_1, \mathbf{C}_1)$ is derived from joint covariance $\mathbf{M}_1$, and $(\mathbf{c}_2, \mathbf{C}_2)$ is derived from joint covariance $\mathbf{M}_2$, the estimate $(\mathbf{c}_1, \mathbf{C}_1)$ will represent the fused estimate if $\det(\mathbf{C}_1) \leqslant \det(\mathbf{C}_2)$.

Although the joint covariance structure underpins the formal analysis of CI (see Appendix A), the actual product of the CI fusion algorithm is just a non-linear mixture of the information contained in the given estimates that are fused. However, the mixture has characteristics that are distinct from those of the Kalman filter. For example, if two estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$ are given, and $\mathbf{A} = \mathbf{B}$, the Kalman filter will exploit assumed statistical independence to produce a fused estimate with covariance $\mathbf{C} = \frac{1}{2}\mathbf{A}$. Because CI does not assume independence, and therefore must be consistent even for the case in which the estimates are completely correlated, its fused estimate has $\mathbf{C} = \mathbf{A}$. This is exactly what is required to ensure consistency in a distributed data fusion context in which a node may receive redundant copies of the same estimate.

In the case of estimates in which $\mathbf{A} < \mathbf{B}$, CI cannot assume any information contribution from the estimate $(\mathbf{b}, \mathbf{B})$, so the fused result is $(\mathbf{a}, \mathbf{A})$. If $\mathbf{A} \not\leqslant \mathbf{B}$ and $\mathbf{B} \not\leqslant \mathbf{A}$, then one estimate has a smaller covariance with respect to some combination of the state variables and the other estimate provides a smaller covariance for a different combination. In this case CI combines information from the two estimates to produce a fused estimate that is superior to either of the given estimates. More specifically, CI by definition determines the consistent joint covariance that yields the smallest possible fused estimate covariance according to the specified measure.

Any consistent joint covariance is sufficient to produce a fused estimate that guarantees consistency, but it is also necessary to guarantee *non-divergence*. A fusion rule is defined to be non-divergent if and only if it never produces a fused estimate that is worse, according to some fixed criterion, than either of the given estimates that were combined to produce it [25]. More generally, the fusion rule should ensure non-divergence over a sequence of fusion operations. For example, if an estimate is maintained of the position of an object of interest, the quality of the estimate should be monotonically non-decreasing as new information is incorporated into it. Non-divergence is achieved with CI by choosing a particular measure, e.g., determinant, that is minimized at each fusion operation. This measure then represents a non-divergence criterion because the size of the estimate covariance according to the selected measure never increases. [5]

[5] Unlike the linear Kalman filter, the non-linear CI fusion operation is not associative. Therefore, the iterative application of CI to process a sequence of estimates is suboptimal with respect to a batch application of CI to simultaneously fuse the set of estimates. However, the iterative application of CI still guarantees consistency and non-divergence. Appendix A provides a formal description and analysis of CI to elucidate its various properties. Appendix B describes generalizations that allow CI to exploit partial known statistical relationships. These generalizations subsume the Kalman filter as a special case. Appendix C addresses common misconceptions about CI.

The notions of consistency and non-divergence are directly pertinent to the decentralized data fusion problem. Specifically, it is possible to maintain consistency within the context of a general distributed network of information processing nodes as long as each node only broadcasts consistent estimates and performs data fusion with implicit joint covariances that are guaranteed to be consistent [13]. However, any practical data fusion system is susceptible to faults that may lead to violations of estimate consistency. Robustness to such faults is necessary to maintain the integrity of the information in the system; otherwise, the occurrence of a single spurious estimate could lead to the progressive corruption of every piece of information in the network.

6. Covariance Union (CU)

CI addresses the general form of the data fusion problem for mean and covariance estimates, but in practice a different problem can arise before data fusion can even be performed. Specifically, what is to be done if two estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$, purportedly relating to the state of the same real-world object, are determined to be mutually inconsistent with each other, i.e., the differences between their means is much larger than what can be expected based on their respective error covariance estimates? For example, if two mean position estimates differ by more than a kilometer, but their respective covariances suggest that each mean is accurate to within a meter, then clearly something is wrong.

One mechanism for detecting statistically significant deviations between estimates is to compute Mahalanobis distances [16,27]. The Mahalanobis distance between estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$ is defined as

$$(\mathbf{a} - \mathbf{b})^{\mathrm{T}}(\mathbf{A} + \mathbf{B})^{-1}(\mathbf{a} - \mathbf{b}), \tag{9}$$

which is essentially just the squared distance between the means as normalized by the sum of their respective covariances. Intuitively, if the covariances are large, then a large difference between the mean vectors $\mathbf{a}$ and $\mathbf{b}$ is not surprising, so the Mahalanobis distance is small. However, if the covariances are very small, then even small differences between the means may yield a large Mahalanobis distance. A large Mahalanobis distance may tend to indicate that the estimates are not consistent with each other, but a user-defined threshold is required to define what constitutes an acceptable deviation. [6] When the threshold is exceeded, the estimates are re-

[6] It must be emphasized that the use of a threshold on Mahalanobis distance is not the only possible mechanism for identifying potentially spurious estimates, but *some* user-defined mechanism is required. Otherwise there is no way to distinguish fault conditions from low probability events. In other words, models for fault conditions are inherently application-specific.

garded as being contradictory and some kind of action must be taken. Resolving such inconsistencies among estimates is referred to as *deconfliction* [18].

The Covariance Intersection method guarantees consistency as long as the estimates to be fused are each consistent. In the deconfliction problem it is only known that one of the estimates, either $(\mathbf{a}, \mathbf{A})$ or $(\mathbf{b}, \mathbf{B})$, is a consistent estimate of the state of the object of interest. Because it is not generally possible to know which estimate is spurious, the only way to rigorously combine the estimates is to form a unioned estimate, $(\mathbf{u}, \mathbf{U})$, that is guaranteed to be consistent with respect to *both* of the two estimates. Such a unioned estimate can be constructed by computing a mean vector $\mathbf{u}$ and covariance matrix $\mathbf{U}$ such that

$$\mathbf{U} \geqslant \mathbf{A} + (\mathbf{u} - \mathbf{a})(\mathbf{u} - \mathbf{a})^{\mathrm{T}}, \tag{10}$$

$$\mathbf{U} \geqslant \mathbf{B} + (\mathbf{u} - \mathbf{b})(\mathbf{u} - \mathbf{b})^{\mathrm{T}}, \tag{11}$$

where some measure of the size of $\mathbf{U}$, e.g., determinant, is minimized. This CU of the two estimates can be subsequently fused with other consistent estimates using CI. (Note that the above definition of CU generalizes immediately to the case of more than two estimates. Further generalizations are discussed in Appendix B.)

Intuitively, the above equations simply say that if the estimate $(\mathbf{a}, \mathbf{A})$ is consistent, then the translation of the vector $\mathbf{a}$ to $\mathbf{u}$ will require its covariance to be enlarged by the addition of a matrix at least as large as the outer product of $(\mathbf{u} - \mathbf{a})$ in order to be consistent. The same reasoning applies if the estimate $(\mathbf{b}, \mathbf{B})$ is consistent. CU therefore determines the mean vector $\mathbf{u}$ having the smallest covariance $\mathbf{U}$ that is large enough to guarantee consistency regardless of which of the two given estimates is consistent. The resulting covariance may be significantly larger than either of the given covariances, but this is an accurate reflection of the actual uncertainty that exists due to the conflict between the two estimates. The key fact is that the CU estimate can be used in any subsequent operation that requires consistency. For example, suppose different emergency response personnel report precise, though mutually inconsistent, estimates of the location to which a rescue helicopter is to be dispatched. The resulting CU combined estimate may have a large covariance that is still sufficient for the helicopter to reach the vicinity of the emergency location.

As a simple example of a CU construction, consider two estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$ of the location of an object observed from two nodes in a network. The estimate from the first node places the mean position at $\mathbf{a} = [0, 0]^{\mathrm{T}}$, and the second node places it at $\mathbf{b} = [4, 4]^{\mathrm{T}}$, and each has an error covariance equal to the identity matrix $\mathbf{I}$. If it is determined that the two estimates are statistically inconsistent with each other, thus implying that one of the estimates is not a consistent estimate of object's location, then deconfliction must be performed. The optimal CU deconflicted estimate can be determined (e.g., by using efficient determinant optimization routines [31]) to be

$$\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}. \tag{12}$$

It is straightforward to verify that this estimate $(\mathbf{u}, \mathbf{U})$ is in fact consistent with respect to either/both of the estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$. If $(\mathbf{a}, \mathbf{A})$ is a consistent estimate of the target's state, then the covariance $\mathbf{U}$ for mean $\mathbf{u}$ must be greater than or equal to $\mathbf{A} + (\mathbf{u} - \mathbf{a})(\mathbf{u} - \mathbf{a})^{\mathrm{T}}$, which it is. It can be verified that the estimate $(\mathbf{u}, \mathbf{U})$ is similarly consistent with respect to the estimate $(\mathbf{b}, \mathbf{B})$. Therefore, if either of the two estimates represents a consistent estimate of the state of the object, then the CU estimate is also consistent.

7. Fault-tolerant network information management

The real power of CU can be demonstrated in an example of a distributed network of information processing nodes. In the most general case, a network may have any connectivity, so no assumptions of acyclicity, fully connectedness, nor any other special structure can be made. In fact, the number of nodes and their connectivity may change dynamically such that the global structure of the network is not known to any node at any time. Consider the situation in which a network has converged to a state in which all nodes contain the same consistent information. More specifically, consider a single estimate $(\mathbf{a}, \mathbf{A})$ that is stored at every node and relates to some particular object of interest. If the estimate in a particular node is replaced with a spurious and inconsistent estimate $(\mathbf{a}^*, \mathbf{A}^*)$, what are the long-term consequences for the consistency of the information in the network as this spurious information propagates?

Suppose the spurious estimate is produced in node i and that it deviates significantly from the true state of the object of interest. In other words, it is possible to conclude that $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{a}^*, \mathbf{A}^*)$ are incompatible because, e.g., their Mahalanobis distance exceeds a chosen threshold. Two cases can be considered: synchronous and asynchronous information exchange.

In the synchronous case, node i and an adjacent node exchange their respective estimates $(\mathbf{a}^*, \mathbf{A}^*)$ and $(\mathbf{a}, \mathbf{A})$. Both nodes independently determine that the estimates are inconsistent with each other. Given that neither node can determine which of the estimates is spurious, both apply CU to produce a consistent estimate $(\mathbf{u}, \mathbf{U})$. At this point the spurious estimate $(\mathbf{a}^*, \mathbf{A}^*)$ has been eliminated from the network and replaced with a consistent estimate. Importantly, consistency has been restored to the network without having to specifically determine which estimate was spurious. The only information cost comes from the fact that the consistent estimate $(\mathbf{a}, \mathbf{A})$ in

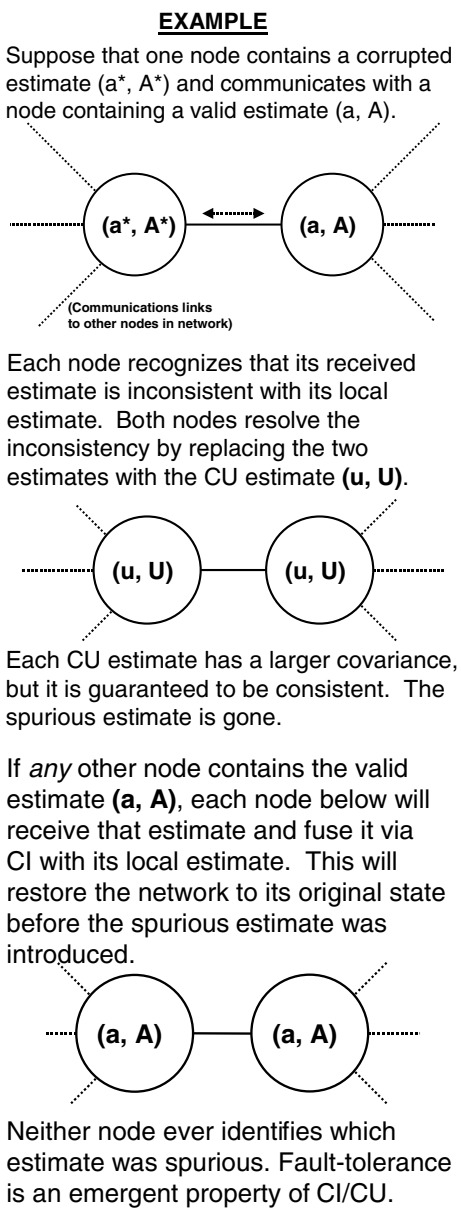

Fig. 1. (a) Two nodes contain mutually inconsistent estimates; (b) CU resolves the inconsistency between estimates and (c) CI incorporates better estimates from other nodes.

the adjacent node has been replaced with an estimate that has a potentially much larger covariance. However, subsequent communications with other nodes will cause, via CI data fusion, the CU estimates to be replaced with the original estimate $(\mathbf{a}, \mathbf{A})$. [7] In other words, the network returns to its original converged state. (See the example in Fig. 1a–c.)

Although the synchronous case provides insight into the workings of the CI/CU information management mechanism, the more general asynchronous case reveals its flexibility. In the asynchronous case there are two ways in which the spurious estimate can be combined with information from other nodes

1. The consistent estimate $(\mathbf{a}, \mathbf{A})$ is transmitted to node i from an adjacent node.
2. The inconsistent estimate $(\mathbf{a}^*, \mathbf{A}^*)$ is transmitted from node i to an adjacent node.

In the first case, the trigger for deconfliction occurs and the two estimates are unioned using CU. The resulting consistent estimate $(\mathbf{u}, \mathbf{U})$ replaces the inconsistent one at node i. At this point the estimate $(\mathbf{a}^*, \mathbf{A}^*)$ is gone from the network, and all estimates in the network are consistent, though not identical. Subsequent receipt of information at node i from other nodes will not trigger deconfliction, and because $\mathbf{A} \leqslant \mathbf{U}$, CI will fuse the estimates $(\mathbf{u}, \mathbf{U})$ and $(\mathbf{a}, \mathbf{A})$ to produce the original estimate $(\mathbf{a}, \mathbf{A})$. This returns the network to exactly the same state that it was in after convergence and before the spurious estimate was introduced.

In the second case, a node adjacent to node i receives the spurious estimate. Deconfliction is triggered and the estimate in the node is replaced with the CU estimate $(\mathbf{u}, \mathbf{U})$. At this point node i still contains the spurious estimate, and the estimate in the adjacent node has been degraded, although it is consistent. Subsequent communications among nodes can produce one of the following situations:

1. If node i receives information from another node (other than the one containing the degraded estimate), then its inconsistent estimate is replaced with $(\mathbf{u}, \mathbf{U})$, and $(\mathbf{a}^*, \mathbf{A}^*)$ is gone from the network. Subsequent CI fusion steps will cause all CU estimates to be replaced with $(\mathbf{a}, \mathbf{A})$, and the network returns to its original state.
2. If the adjacent node receives another transmission from node i, then deconfliction will not be triggered because $(\mathbf{a}^*, \mathbf{A}^*)$ and $(\mathbf{u}, \mathbf{U})$ are not incompatible estimates. The consequence is that CI will be applied to fuse the two estimates, and the adjacent node will now store the spurious estimate $(\mathbf{a}^*, \mathbf{A}^*)$. Note that this can only occur if node i transmits its estimate twice to the adjacent node before either node receives information from any other node in the network. At this point, the spurious estimate has proliferated to another node.

The key observation to be made is that most sequences of communications result in the complete elimination of the spurious estimate. However, it is possible to construct a sequence of communications among nodes that causes the spurious estimate to propagate from one node to another. This propagation

[7] This is because, as discussed in the previous section, if CI is given estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$, $\mathbf{A} < \mathbf{B}$, the fused estimate will be $(\mathbf{a}, \mathbf{A})$. The CU estimate satisfies $\mathbf{U} > \mathbf{A}$, so the CI fused estimate is $(\mathbf{a}, \mathbf{A})$.

is tied strongly to the relative number of spurious and non-spurious estimates transmitted in the network. Specifically, every time the spurious estimate is transmitted to a node containing the estimate $(\mathbf{a}, \mathbf{A})$, the latter estimate is replaced with a CU estimate. The relative likelihood of that CU estimate being fused via CI with a consistent versus an inconsistent estimate is then directly related to the relative numbers of consistent and inconsistent estimates in the network. If the majority of the estimates are consistent, then one can expect the network to ultimately converge to its original state. The behavior of the network can only be analyzed probabilistically because of the stochastic nature of unrestricted asynchronous communications among nodes. Absolute convergence can only be guaranteed under more restrictive modes of network communications.

The above example assumes that it is possible to recognize that the spurious estimate is inconsistent with current estimates in the network so that deconfliction is triggered and CU is applied. This is typically the case when the spurious estimate is produced by a sensor failure or human data entry error, but what happens if an adversary exploits knowledge of the estimates in the network and constructs an inconsistent estimate that falls just below the deconfliction threshold? In such a case the spurious estimate will be fused with estimates in the network, thus pulling them toward inconsistency. However, the deconfliction threshold imposes a strong limit on the deviation between the spurious and real estimates, so the absolute impact of a spurious estimate on the information in the network is strongly limited. The only way an adversary can completely corrupt the information in the network is to continually introduce many spurious estimates at points throughout the network to progressively drive all estimates further and further away from ground truth.

It is a remarkable fact that the combination of CI and CU can, in many cases, completely eliminate spurious pieces of information that may be introduced as the result of intentional or unintentional data corruption. This is important because as a network grows, the number of points of failure increases. Also important is the fact that the CI/CU mechanism does not exploit anything beyond the information local to a node. When a node is faced with the deconfliction of the estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{a}^*, \mathbf{A}^*)$, the conflict is ultimately resolved without having to, e.g., resort to a global examination of information in the network to try to ascertain which estimate is spurious. The combination of CI and CU therefore offers truly decentralized fault tolerance.

8. Discussion

What CU provides is a mathematically rigorous mechanism for replacing two or more mutually inconsistent mean and covariance estimates with a single provably consistent one. In order to guarantee that this combined estimate is consistent with the actual state of the object of interest, its error covariance must be large enough to compensate for the amount of deviation that exists between the two given estimates. If there is a large deviation, then the covariance of the combined estimate will be relatively much larger than the covariances of the given estimates. However, if there is no way to determine which of the two estimates conservatively reflects the actual state of the object of interest, then a large covariance reflects the true degree of uncertainty that exists.

The combination of CU and CI provides a consistent and fault-tolerant solution for the network information management problems that arise in decentralized command and control applications. Specifically, CI ensures that the fusion of consistent estimates from different nodes produces consistent fused estimates under all circumstances without any restrictions placed on the topology of the network. CU complements CI by offering a mechanism for handling situations in which inconsistent estimates enter the network. The combination of CI and CU causes estimates that are inconsistent with the consensus of estimates in the network to be either filtered out or at least limited in their ability to proliferate. This provides robustness to a data fusion network that could otherwise be undermined by the introduction of even a single spurious estimate. Most importantly, this robustness is an emergent property of the local application of CI and CU operations at each node.

In summary, the CU theory provides a general and mathematically rigorous mechanism for estimate deconfliction that can be applied directly in general decentralized data fusion systems. However, the CU approach is likely to find uses in a broader range of applications in which mean and covariance estimates are maintained. Examples include almost any current application of the Kalman filter or CI, e.g., for control or tracking. For example, CU can be used to augment multi-hypothesis methods for dealing with assignment ambiguity when system resource limits have been reached in multiple-target tracking applications. In this context it offers an alternative to assumed probability distribution-based methods [15]. CU can also be used to produce consistent estimates for predictions from multiple models when the identity of the correct model is not known. More generally, CU and CI can be applied to complement each other in a variety of applications for which the conditions required by the Kalman filter cannot be satisfied.

Acknowledgements

The author would like to acknowledge the support of IDAK Industries and continuing gratitude to Mike Brady of Oxford University.

Appendix A. Consistency and tightness of CI

Although the CI fusion method is generally defined in Section 5, this appendix provides formal analyses of its properties and special structures. To simplify the exposition, all estimates are assumed to be in the same state space, but standard methods can be applied to accommodate estimates in different state spaces when the projection from one state space to the other is known [2,17]. It should also be noted that the term *positive definite* will be used to refer to matrices that may also be positive *semidefinite*. In some cases algebraic manipulations may implicitly assume strict positive definiteness when taking inverses, but almost all results in this appendix can be derived more generally to accommodate singular matrices.

One way to think about the general mean and covariance data fusion problem is to explicitly represent the joint covariance structure that implicitly exists between a given pair of estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$

$$\begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix}, \tag{A.1}$$

where $\mathbf{X}$ represents the actual, but unknown, cross covariance between the two estimates. If $\mathbf{X}$ were known, then it would be possible to apply more general formulations of the Kalman filter equations to produce an optimal fused estimate. Unfortunately, these generalizations only guarantee consistency if the cross covariance is known *exactly*, i.e., it cannot be conservatively approximated in any way analogous to the way conservative covariance estimates are used.

Without knowledge of $\mathbf{X}$, the only way to ensure consistency in the application of the Kalman filter is to identify a joint covariance that is guaranteed to be consistent based on the information available. In the present context, therefore, a joint covariance matrix $\mathbf{M}$ must be determined such that

$$\mathbf{M} \geqslant \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix} \tag{A.2}$$

for every possible cross covariance $\mathbf{X}$. It can be inferred from the symmetry of the unknown cross covariance information (i.e., $\mathbf{M}$ must be consistent for any instantiation $\mathbf{X} = \mathscr{X}$ and for $\mathbf{X} = -\mathscr{X}$) that the off-diagonal blocks of $\mathbf{M}$ should be zero, and its diagonal blocks must be sufficiently larger than $\mathbf{A}$ and $\mathbf{B}$ to account for the effects of all possible degrees of correlation among the error components of the mean estimates $\mathbf{a}$ and $\mathbf{b}$.

A consistent and tight joint covariance $\mathbf{M}$ can be generated by selecting a scalar value ω, $0 \leqslant \omega \leqslant 1$, and verifying that

$$\begin{bmatrix} \frac{1}{\omega}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\omega)}\mathbf{B} \end{bmatrix} \geqslant \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix}. \tag{A.3}$$

An examination of the difference between the ω-parameterized covariance, which will be referred to as the CI joint covariance, and the unknown joint covariance yields the following. [8]

$$\begin{bmatrix} \frac{1}{\omega}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\omega)}\mathbf{B} \end{bmatrix} - \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{1}{\omega} - 1\right)\mathbf{A} & -\mathbf{X} \\ -\mathbf{X}^{\mathrm{T}} & \left(\frac{1}{(1-\omega)} - 1\right)\mathbf{B} \end{bmatrix} \tag{A.4}$$

$$= \begin{bmatrix} \frac{(1-\omega)}{\omega}\mathbf{A} & -\mathbf{X} \\ -\mathbf{X}^{\mathrm{T}} & \frac{\omega}{(1-\omega)}\mathbf{B} \end{bmatrix}. \tag{A.5}$$

In order to prove consistency, the above difference must be positive semidefinite. This condition can be established using the contraction criterion [10], which states that for any positive semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$ with cross covariance $\mathbf{X}$, the partitioned matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^{\mathrm{T}} & \mathbf{B} \end{bmatrix} \tag{A.6}$$

is positive semidefinite if and only if there is a contraction matrix $\mathbf{\Omega}$ such that:

$$\mathbf{X} = \mathbf{A}^{1/2}\mathbf{\Omega}\mathbf{B}^{T/2}, \tag{A.7}$$

where a contraction matrix is simply a matrix whose largest singular value is less than or equal to unity. Therefore, if the partitioned matrix represents the joint covariance with unknown $\mathbf{X}$, then the above equation defines the set of admissable instantiations of $\mathbf{X}$.

Applying the contraction criterion to the difference matrix in Eq. (A.5) reveals the following:

$$\mathbf{X} = -\left[\frac{(1-\omega)}{\omega}\mathbf{A}\right]^{1/2}\mathbf{\Omega}\left[\frac{\omega}{(1-\omega)}\mathbf{B}\right]^{1/2} \tag{A.8}$$

$$= -\left[\frac{(1-\omega)\omega}{\omega(1-\omega)}\right]^{1/2}\mathbf{A}^{1/2}\mathbf{\Omega}\mathbf{B}^{1/2} \tag{A.9}$$

$$= \mathbf{A}^{1/2}\mathbf{\Omega}'\mathbf{B}^{1/2}, \tag{A.10}$$

where $\mathbf{\Omega}'$ is just a re-labeling of the set of all possible contraction matrices (the singular values of a contraction are invariant with respect to changes in sign). Thus, the condition for the difference matrix of Eq. (A.5) to be positive semidefinite is equivalent to the condition required for the original joint covariance to be positive semidefinite. In other words, the CI joint covariance is consistent with respect to the true unknown joint covariance if and only if the unknown joint covariance is positive definite. This establishes biconditional tightness for the ω-parameterized family of joint covariances.

The above CI result based on the contraction criterion can be applied to deduce solutions to variety of

[8] This analysis of CI follows the original tightness and consistency analysis in Appendix 14 of [25].

related problems. For example, some filtering applications (e.g., weather modeling [24], simultaneous localization and map building [25], and other large state space problems) cannot afford the computational resources necessary to maintain large covariance matrices. It is therefore necessary to approximate a full covariance (or submatrix thereof) with one that is block diagonal, e.g.,

$$\begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \geqslant \begin{bmatrix} \mathbf{A} & \mathbf{Y} \\ \mathbf{Y}^{\mathrm{T}} & \mathbf{B} \end{bmatrix}, \tag{A.11}$$

when the cross covariance $\mathbf{Y}$ is known. Using the contraction criterion, it is possible to deduce

$$\begin{bmatrix} \mathbf{A}+\mathbf{Y}\mathbf{S}^{-1}\mathbf{Y}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}+\mathbf{S} \end{bmatrix} - \begin{bmatrix} \mathbf{A} & \mathbf{Y} \\ \mathbf{Y}^{\mathrm{T}} & \mathbf{B} \end{bmatrix} \geqslant \mathbf{0}. \tag{A.12}$$

This can be verified by applying the criterion to the difference matrix

$$\begin{bmatrix} \mathbf{Y}\mathbf{S}^{-1}\mathbf{Y}^{\mathrm{T}} & -\mathbf{Y} \\ -\mathbf{Y}^{\mathrm{T}} & \mathbf{S} \end{bmatrix}, \tag{A.13}$$

which for $\mathbf{\Omega} = -\mathbf{I}$, satisfies [9]

$$-\mathbf{Y} = \mathbf{Y}\mathbf{S}^{-(1/2)}\mathbf{\Omega}\mathbf{S}^{1/2} \tag{A.14}$$

$$= -\mathbf{Y}. \tag{A.15}$$

Thus, the condition holds for all $\mathbf{S}$, where $\mathbf{S}$ can be computed to minimize any chosen measure of covariance size. [10] The family of upper bound covariances can also be characterized in the form

$$\begin{bmatrix} \mathbf{A}+\mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}+\mathbf{Y}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{Y} \end{bmatrix}, \tag{A.16}$$

which is required when the dimensionality of $\mathbf{B}$ is greater than that of $\mathbf{A}$. It should be noted that the above results can be applied to eliminate any subset of non-zero cross covariance terms in the approximation of a general covariance matrix by applying (implicitly) permutation matrices to group the designated terms into a single block. The same procedure can also be repeated to eliminate multiple blocks.

The formulation of Eq. (A.3) can be extended to apply to any number of estimates. For example, a joint covariance can been generated for estimates $(\mathbf{a}_1, \mathbf{A}_1)$, $(\mathbf{a}_2, \mathbf{A}_2), \ldots, (\mathbf{a}_n, \mathbf{A}_n)$ in inverse form as

$$\begin{bmatrix} \omega_1\mathbf{A}_1^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \omega_2\mathbf{A}_2^{-1} & \mathbf{0} & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \omega_n\mathbf{A}_n^{-1} \end{bmatrix}. \tag{A.17}$$

A consistent augmented joint covariance that includes an estimate $(\mathbf{a}_{n+1}, \mathbf{A}_{n+1})$ can be constructed as

$$\begin{bmatrix} (1-\omega_{n+1}) \begin{bmatrix} \omega_1\mathbf{A}_1^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \omega_2\mathbf{A}_2^{-1} & \mathbf{0} & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \omega_n\mathbf{A}_n^{-1} \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \omega_{n+1}\mathbf{A}_{n+1}^{-1} \end{bmatrix}. \tag{A.18}$$

Because $1 = \sum_{i=1}^{n} \omega_i$, it must be the case that $1 = \omega_{n+1} + (1-\omega_{n+1})\sum_{i=1}^{n} \omega_i$, so the weights sum appropriately. However, there is no constraint that the relative weightings given by $\omega_1, \ldots, \omega_n$ must be maintained in the computation of ω_{n+1}. The sequential pairwise CI fusion of a maintained estimate with each newly obtained estimate implicitly assumes such a constraint. The augmented joint covariance that yields the best possible fused estimate must be computed by jointly optimizing over the $n+1$ parameters $\omega_1, \ldots, \omega_{n+1}$, which is the reason why, unlike the Kalman filter, the CI fusion operation is non-associative.

The CI fusion equations for n estimates with completely unknown degrees of correlation (cross covariance) can be obtained from the inverse Kalman filter equations applied to the inverse CI joint covariance

$$\mathbf{C} = \left(\omega_1\mathbf{H}_1^{\mathrm{T}}\mathbf{A}_1^{-1}\mathbf{H}_1 + \omega_2\mathbf{H}_2^{\mathrm{T}}\mathbf{A}_2^{-1}\mathbf{H}_2 + \cdots + \omega_n\mathbf{H}_n^{\mathrm{T}}\mathbf{A}_n^{-1}\mathbf{H}_n\right)^{-1}, \tag{A.19}$$

$$\mathbf{c} = \mathbf{C}\left(\omega_1\mathbf{H}_1^{\mathrm{T}}\mathbf{A}_1^{-1}\mathbf{a}_1 + \omega_2\mathbf{H}_2^{\mathrm{T}}\mathbf{A}_2^{-1}\mathbf{a}_2 + \cdots + \omega_n\mathbf{H}_n^{\mathrm{T}}\mathbf{A}_n^{-1}\mathbf{a}_n\right), \tag{A.20}$$

where $\mathbf{H}_i$ is the transformation from the state space of the fused estimate to the state space of estimate i (see [19] for the form of the fused mean when $\mathbf{H}_i$ is a linearized approximation to a non-linear transformation). The ω values can be computed to maximize the determinant of the inverse covariance (hence minimizing the determinant of the covariance) using a variety of standard convex and semidefinite programming packages [4,31]. In the case of two estimates, optimizing the single variable ω is particularly efficient after the covariances have been simultaneously diagonalized.

In contexts in which all past information is summarized in the form of a single mean and covariance estimate, the sequential application of CI maximally exploits all information available at each update, but a better (smaller covariance) summary of the information in a sequence of estimates can be generated from the

[9] Because the actual cross covariance is given ($\mathbf{Y}$ is not a variable), all that is needed to verify consistency is to identify a particular contraction matrix that satisfies the criterion. This contrasts with the CI problem in which $\mathbf{X}$ parameterizes a family of joint covariances, the true value of which is unknown.

[10] Simon Julier has applied SVD methods to obtain a minimum trace solution directly (personal correspondence).

batch fusion of the past estimates if they are available. In either context, the application of CI (with a fixed measure of the size of the fused covariance at each application) guarantees consistency and non-divergence for any sequence of consistent mean and covariance estimates.

Appendix B. Split covariance estimates

CI can be applied in a more general context to exploit partial information about the degree of correlation that exists between a given estimate and other estimates with which it may be fused. Specifically, the CI equations can be generalized to accommodate estimates maintained in *split covariance* form [12]. A split covariance is one in which a consistent error covariance matrix can be determined for errors in the mean that are known to be statistically independent of errors in other estimates. This known independence can be exploited in the Split CI equations to provide an optimal hybrid of CI and Kalman. In order to apply CU with Split CI, it is necessary to generalize CU to be applied with split covariance estimates.

B.1. Split Covariance Intersection

CI and the Kalman filter represent opposite extremes with respect to their assumptions about two given estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$ when no other information is available about their statistical relationship. The Kalman filter implicitly assumes that the two estimates are uncorrelated while CI implicitly assumes that the two estimates may have any possible cross covariance, e.g., they may be completely correlated or anti-correlated. Consequently, the Kalman filter tends to underestimate the covariance of the error in its fused estimates while CI tends to overestimate the covariance. In practice, however, it is common to encounter mixtures of information in which some statistical relationships are known and some are not.

The CI joint covariance for estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$ can be expressed (see Appendix A) as

$$\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \frac{1}{\omega}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\omega)}\mathbf{B} \end{bmatrix} \right), \tag{B.1}$$

which is guaranteed to be consistent for whatever true, but unknown, cross covariance exists between the errors in the two estimates. Now suppose that $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$ and $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$, where $\mathbf{a}_1$ and $\mathbf{b}_1$ are correlated to an unknown degree, while the errors associated with $\mathbf{a}_2$ and $\mathbf{b}_2$ are assumed to be independent. Also, let the respective covariances of the components be $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{B}_1$, and $\mathbf{B}_2$. From the above results we can form a consistent joint system can be formed as

$$\left(\begin{bmatrix} \mathbf{a}_1 + \mathbf{a}_2 \\ \mathbf{b}_1 + \mathbf{b}_2 \end{bmatrix}, \begin{bmatrix} \frac{1}{\omega}\mathbf{A}_1 + \mathbf{A}_2 & \mathbf{0} \\ \mathbf{0} & \frac{1}{1-\omega}\mathbf{B}_1 + \mathbf{B}_2 \end{bmatrix} \right), \tag{B.2}$$

where the covariance is just the sum of

$$\begin{bmatrix} \frac{1}{\omega}\mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{1-\omega}\mathbf{B}_1 \end{bmatrix} + \begin{bmatrix} \mathbf{A}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}. \tag{B.3}$$

A byproduct of the above derivation is that $\frac{1}{\omega}\mathbf{A} + \frac{1}{1-\omega}\mathbf{B}$ represents a consistent covariance for $\mathbf{a} + \mathbf{b}$ when $\mathbf{a}$ and $\mathbf{b}$ are correlated to an unknown extent. This operation is referred to as *Covariance Addition* (CA). [11]

Letting $\mathbf{A} = \frac{1}{\omega}\mathbf{A}_1 + \mathbf{A}_2$ and $\mathbf{B} = \frac{1}{1-\omega}\mathbf{B}_1 + \mathbf{B}_2$, the following generalized CI equations can be obtained from the inverse Kalman filter equations (see Eqs. (3) and (4)):

$$\begin{aligned} \mathbf{C} &= [\mathbf{A}^{-1} + \mathbf{B}^{-1}]^{-1} \\ &= \left[\left(\frac{1}{\omega}\mathbf{A}_1 + \mathbf{A}_2 \right)^{-1} + \left(\frac{1}{1-\omega}\mathbf{B}_1 + \mathbf{B}_2 \right)^{-1} \right]^{-1}, \end{aligned} \tag{B.4}$$

$$\begin{aligned} \mathbf{c} &= \left[\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b} \right] \\ &= \mathbf{C}\left[\left(\frac{1}{\omega}\mathbf{A}_1 + \mathbf{A}_2 \right)^{-1}\mathbf{a} + \left(\frac{1}{1-\omega}\mathbf{B}_1 + \mathbf{B}_2 \right)^{-1}\mathbf{b} \right], \end{aligned} \tag{B.5}$$

where the known independence of the errors associated with $\mathbf{a}_2$ and $\mathbf{b}_2$ is exploited.

Although the above generalization of CI exploits available knowledge about independent error components, further exploitation is impossible because the combined covariance $\mathbf{C}$ is formed from *both* independent and correlated error components. However, CI can be generalized to maintain separate covariance components, $\mathbf{C}_1$ and $\mathbf{C}_2$, reflecting the correlated and known-independent error components, respectively. This generalization is referred to as *Split CI*.

Letting $\tilde{\mathbf{a}}_1$ and $\tilde{\mathbf{a}}_2$ be the correlated and known-independent error components of $\mathbf{a}$, with $\tilde{\mathbf{b}}_1$ and $\tilde{\mathbf{b}}_2$ similarly defined for $\mathbf{b}$, the errors $\tilde{\mathbf{c}}_1$ and $\tilde{\mathbf{c}}_2$ can be expressed in information (inverse covariance) form as

$$\mathbf{C}^{-1}(\tilde{\mathbf{c}}_1 + \tilde{\mathbf{c}}_2) = \mathbf{A}^{-1}(\tilde{\mathbf{a}}_1 + \tilde{\mathbf{a}}_2) + \mathbf{B}^{-1}(\tilde{\mathbf{b}}_1 + \tilde{\mathbf{b}}_2), \tag{B.6}$$

from which the following can be obtained after pre-multiplying by $\mathbf{C}$

$$(\tilde{\mathbf{c}}_1 + \tilde{\mathbf{c}}_2) = \mathbf{C}\left[\mathbf{A}^{-1}(\tilde{\mathbf{a}}_1 + \tilde{\mathbf{a}}_2) + \mathbf{B}^{-1}(\tilde{\mathbf{b}}_1 + \tilde{\mathbf{b}}_2) \right]. \tag{B.7}$$

Squaring both sides, taking expectations, collecting independent terms, and letting $\mathbf{A} = \frac{1}{\omega}\mathbf{A}_1 + \mathbf{A}_2$ and $\mathbf{B} = \frac{1}{1-\omega}\mathbf{B}_1 + \mathbf{B}_2$ yields

[11] Note that the CA result can be used to construct an alternative covariance normalizing term in the Mahalanobis distance expression when comparing potentially correlated estimates.

$$\mathbf{C}_2 = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}(\mathbf{A}^{-1}\mathbf{A}_2\mathbf{A}^{-1} + \mathbf{B}^{-1}\mathbf{B}_2\mathbf{B}^{-1}) \times (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}, \quad \text{(B.8)}$$

where the non-independent part can be obtained simply by subtracting the above result from the overall fused covariance $\mathbf{C} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}$, in other words,

$$\mathbf{C}_1 = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} - \mathbf{C}_2. \quad \text{(B.9)}$$

The above analysis applies to two estimates, but it can also be applied to generate a batch estimate that optimally combines more than two estimates. *Split CA* equations can be similarly derived. These results can be further generalized to include known cross covariances, e.g.,

$$\left(\begin{bmatrix}\mathbf{a}_1 + \mathbf{a}_2 \\ \mathbf{b}_1 + \mathbf{b}_2\end{bmatrix}, \begin{bmatrix}\frac{1}{\omega}\mathbf{A}_1 + \mathbf{A}_2 & \mathbf{Y} \\ \mathbf{Y}^{\mathrm{T}} & \frac{1}{1-\omega}\mathbf{B}_1 + \mathbf{B}_2\end{bmatrix}\right), \quad \text{(B.10)}$$

where the covariance is the sum of

$$\begin{bmatrix}\frac{1}{\omega}\mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{1-\omega}\mathbf{B}_1\end{bmatrix} + \begin{bmatrix}\mathbf{A}_2 & \mathbf{Y} \\ \mathbf{Y}^{\mathrm{T}} & \mathbf{B}_2\end{bmatrix} \quad \text{(B.11)}$$

in which $\mathbf{Y}$ is the known cross covariance between $\tilde{\mathbf{a}}_2$ and $\tilde{\mathbf{b}}_2$. Alternatively, the result of Eq. (A.12) can be applied to maintain block diagonality by eliminating $\mathbf{Y}$.

B.2. Split Covariance Union

Let the estimate $(\mathbf{a}, \mathbf{A})$ be represented in split covariance form $(\mathbf{a}, \mathbf{A}_1, \mathbf{A}_2)$, where $\mathbf{A}_2$ is the covariance of the known independent errors in $\mathbf{a}$ and $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$. The generalized CU problem is to union this estimate with an estimate $(\mathbf{b}, \mathbf{B}_1, \mathbf{B}_2)$ to produce a split CU estimate $(\mathbf{u}, \mathbf{U}_1, \mathbf{U}_2)$. This can be achieved as follows: the standard CU equations are applied to the non-split forms of the estimates to produce the non-split CU estimate $(\mathbf{u}, \mathbf{U})$. The independent component of the CU, covariance $\mathbf{U}_2$, can then be determined as the *largest* covariance matrix satisfying the following inequalities:

$$\mathbf{U}_2 \leqslant \mathbf{A}_2, \quad \text{(B.12)}$$

$$\mathbf{U}_2 \leqslant \mathbf{B}_2. \quad \text{(B.13)}$$

In words, the above inequalities say that independent errors exist with a covariance that is greater than $\mathbf{U}_2$ regardless of which estimate is consistent and which is spurious. Because this statistical independence can be exploited by the Split CI equations to produce better fused estimates, there is benefit to be gained by determining the largest possible consistent estimate of $\mathbf{U}_2$, e.g., with the largest determinant. However, when CU is applied in conjunction with CI, the best possible estimate can be obtained by *jointly* optimizing $\mathbf{U}_2$ along with the ω parameters in the CI equations to minimize the resulting covariance. For example, $\mathbf{U}_2$ can be determined so as to yield the smallest covariance in a subsequent CI step; or alternatively, the ω values can be selected during a CI step to minimize the covariance $\mathbf{U}$ of a subsequent CU step. In general, the best possible estimates are obtained by jointly optimizing as many CI, CA, and CU parameters as possible to exploit as much information as is available. [12]

It should also be noted that under some circumstances the outer product difference in the means can be included in the independent component of the split CU estimate. The determination must be made based on how the estimate subsequently will be used, e.g., the difference between the means for a pair of observed objects will certainly be correlated with the difference at a subsequent observation.

Appendix C. The CI framework

This appendix briefly addresses some common misconceptions about the CI approach to data fusion and related problems. This requires first summarizing the principal contributions of the CI framework for data fusion, which are [25]:

1. *A rigorous notion of estimate consistency*. CI is founded on the definition of a *consistent* estimate $(\mathbf{a}, \mathbf{A})$ as being one in which $\mathbf{A} \geqslant E[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\mathrm{T}}]$, and CI guarantees that the fusion of consistent estimates always yields consistent estimates. Most traditional filter/fusion algorithms have been founded on Bayesian or other probabilistic principles that demand much more restrictive conditions and assumptions, e.g., that the mean vector $\mathbf{a}$ is an unbiased estimate of the conditional mean. The consistency criterion represents a rigorous alternative principle that offers much greater algorithm design flexibility.
2. *A rigorous notion of filter non-divergence*. The term *convergence* has been used traditionally to describe the performance of a static filter/fusion algorithm as it processes a sequence of mean and covariance estimates. Specifically, the estimates produced by the algorithm are said to converge if they monotonically approach some limiting estimate (steady state). CI defines a more relaxed property, *non-divergence*, that requires only that the algorithm's estimates are non-increasing with respect to some fixed measure of covariance size. [13] Designing an algorithm that is provably non-divergent is considerably more tractable

[12] Additional optimization parameters may be introduced if block diagonality is maintained using the result of Eq. (A.12).

[13] Note that the use of the term *divergence* here relates to estimate degradation due to an intrinsic defect in a filtering/fusion algorithm. This is different from application-specific degradation of estimates resulting from the accumulation of data-dependent (e.g., over time) uncertainty.

in many applications than trying to establish more restrictive performance guarantees.

Because there are very few analogs between CI and the Bayesian treatments of the Kalman filter presented in most textbooks, there is a natural inclination to try to interpret CI from a Bayesian perspective. Whether or not this is possible, it is reasonable to consider whether anything can be achieved by doing so.

The original derivation of the Kalman filter was based on orthogonal projection theory, and the fact that there exists a simple Bayesian interpretation of the result when error distributions are Gaussian was presented only as an interesting sidelight [14]. However, many subsequent references on the Kalman filter only present the Bayesian interpretation, and a consequence is that many engineers believe that the Kalman filter *requires* assumptions of Gaussianity. [14] Given that the Bayesian interpretation of the Kalman filter imposes many more restrictions, it is reasonable to question what is gained from such an interpretation. A reliable filter/fusion algorithm should have rigorous theoretical performance guarantees under assumptions that can be satisfied in practice, otherwise there is no way to relate theoretical analysis to practical expected performance. CI's consistency guarantee is rigorous under assumptions that are easily satisfied in practice.

A common misconception about CI arises from attempts to compare it to the Kalman filter. CI was developed to address problems in which the Kalman independence (or known cross covariance) assumptions cannot be satisfied. In other words, CI and the Kalman filter are optimal for disjoint classes of problems, so there is no basis for comparing their relative performance. However, it is sometimes suggested [6] that CI is "too conservative" when compared to the Kalman filter. Such a statement can only be interpreted under conditions that fail to satisfy the assumptions of one or both of the filters. For example, consider the fusion of two given estimates under the following circumstances:

1. *The two estimates are independent*. In this case the Kalman filter produces an optimal fused estimate while CI produces a consistent, though suboptimal, estimate.
2. *The two estimates are completely correlated*. In this case CI produces an optimal fused estimate while Kalman produces an inconsistent estimate.
3. *The two estimates are partially correlated*. In this case CI produces a consistent, though suboptimal, estimate while the Kalman filter produces an inconsistent estimate.

CI produces consistent estimates in all three cases, and is optimal in one case. The Kalman filter is optimal in one case and inconsistent in the others. Therefore it is difficult to make a generic claim that CI is "too conservative." As discussed in Appendix B, however, CI can be generalized to subsume the Kalman filter for accommodating combinations of independent error processes and error processes with unknown degrees of correlation. When complete independence can be assumed, split CI is equivalent to the Kalman filter; and its performance is equivalent to standard CI when no cross covariance information is available. In summary, CI and its various generalizations provide optimal solutions to a broad range of data fusion and filtering problems for mean and covariance estimates.

[14] In fact, one commonly cited motivation for investigating the use of neural networks and similar methods in traditional filtering and control applications is the claim that the Kalman filter imposes restrictive Gaussianity assumptions that often cannot be satisfied.

References

[1] Y. Bar-Shalom, T.E. Fortmann, Tracking and Data Association, The Academic Press, New York, 1988.
[2] S. Blackman, Multiple-target Tracking with Radar Applications, Artech House, Dedham, MA, 1986.
[3] C.L. Bowman, A. Steinberg, A systems engineering approach for implementing data fusion systems, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 16).
[4] S. Boyd, S. Wu, SDPSOL: User's Guide, November 1995. SDPSOL: A Parser/Solver for Semidefinite Programs with Matrix Structure.
[5] R.S. Bucy, K.D. Senne, Digital synthesis of non-linear filters, Automatica 7 (3) (1971) 287–298.
[6] F. Daum, Review of handbook of multisensor data fusion, IEEE Aerospace Electron. Syst. Mag. 16 (10) (2001) 10.
[7] D.L. Hall, J. Llinas, Multisensor data fusion, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 1).
[8] D.L. Hall, A.N. Steinberg, Dirty secrets in multisensor data fusion, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 21).
[9] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.
[10] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991.
[11] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.
[12] S. Julier, J. Uhlmann, General decentralized data fusion with covariance intersection, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 12).
[13] S.J. Julier, J.K. Uhlmann, A non-divergent estimation algorithm in the presence of unknown correlations, in: American Control Conference, Albuquerque, New Mexico, 1997.
[14] R.E. Kalman, A new approach to linear filtering and prediction problems, ASME J. Basic Eng. 82 (1960) 34–45.
[15] T. Kurubarajan, Y. Bar-Shalom, Target tracking using probabilistic data association-based techniques with applications to sonar, radar, and eo sensors, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 8).
[16] P. Mahalanobis, On the generalized distance in statistics, Proc. Natl. Inst. Sci. 12 (1936) 49–55.
[17] P.S. Maybeck, in: Stochastic Models, Estimation, and Control, vol. 1, Academic Press, New York, 1979.

[18] R. Mittu, F. Segaria, Common operational picture (cop) and common tactical picture (ctp) management via a consistent networked information stream (cnis), in: Proceedings of the Command and Control Research and Technology Symposium, 2000.
[19] A.G.O. Mutambara, Decentralized Estimation and Control for Nonlinear Systems, CRC Press, Boca Raton, FL, 1998.
[20] A.B. Poore, S. Sihua Lu, B.J. Suchomel, Data association using multiple frame assignments, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 11).
[21] D.B. Reid, An algorithm for tracking multiple targets, IEEE Trans. Autom. Control AC-24 (6) (1979).
[22] P.A. Smyton, Layered information correlation and aggregation architecture, in: The Proceedings of the 1999 Command and Control Research and Technology Symposium, Newport, RI, 1999.
[23] L.D. Stone, C.A. Barlow, T.L. Corwin, Bayesian Mutiple Target Tracking, Artech House, 1999.
[24] R. Todling, S.E. Cohn, Suboptimal schemes for atmospheric data assimilation based on the Kalman filter, Mon. Weather Rev. 122 (1994) 2530–2557.
[25] J.K. Uhlmann, Dynamic map building and localization: new theoretical foundations, Ph.D. Thesis, University of Oxford, 1995.
[26] J.K. Uhlmann, Algorithms for multiple-target tracking, Am. Scient. 80 (2) (1992).
[27] J.K. Uhlmann, Introduction to the algorithmics of data association in multiple-target tracking, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 3).
[28] J.K. Uhlmann, An introduction to the combinatorics of optimal and approximate data association, in: Handbook of Multisensor Data Fusion, CRC Press, Boca Raton, FL, 2001 (Chapter 9).
[29] S.W. Utete, Network management in decentralised sensing systems, Ph.D. Thesis, Robotics Research Group, Department of Engineering Science, University of Oxford, January 1995.
[30] A. Warren, Trajectory prediction concepts for next generation air traffic management, in: Proceedings of the Air Traffic Management (ATM) 2000 Conference, Napoli, FAA/Eurocontrol, 2000.
[31] S.-P. Wu, L. Vandenberghe, S. Boyd, Maxdet: Software for determinant maximization problems, alpha version, Stanford University Technical Report, April 1996.