Covariance Intersection for Partially Correlated Random Vectors

Zongze Wu, Qianqian Cai, and Minyue Fu, *Fellow, IEEE*

***Abstract*—This paper generalizes the well-known covariance intersection algorithm for distributed estimation and information fusion of random vectors. Our focus will be on partially correlated random vectors. This is motivated by the restriction of the standard covariance intersection algorithm, which treats all random vectors with arbitrary cross correlations and the restriction of the classical Kalman filter, which requires complete knowledge of the cross correlations. We first give a result to characterize the conservatism of the standard covariance intersection algorithm. We then generalize the covariance intersection algorithm to two random vectors with a given correlation coefficient bound and show in what sense the resulting covariance bound is tight. Finally, we generalize the notion of correlation coefficient bound to multiple random vectors and provide a covariance intersection algorithm for this general case. Our results will make the already popular covariance intersection more applicable and more accurate for distributed estimation and information fusion problems.**

***Index Terms*—Covariance intersection (CI), distributed estimation, multisensor data fusion, multisensor information fusion.**

I. Introduction

With the rapid development and wide deployment of low-cost sensors and sensor networks, the problems of distributed estimation and information fusion (or data fusion) have become an important research topic in estimation theory [1]. Multisensor information fusion and network-based distributed estimation go hand by hand, and they form an emerging technology which finds wide applications in surveillance, remote sensing, autonomous vehicles, monitoring of complex machinery, medical diagnosis, robotics, video and image processing, vehicle localization, diffusion Kalman filtering, distributed estimation, and NASA Mars rover [2]–[9], [11]–[17], see [18] for a recent survey on multisensor fusion. As stated in [18], "the essence of multisensor fusion techniques is to combine data from multiple sensors and related information from associated databases, to achieve improved accuracies and more specific inferences than that could be achieved by the use of a single sensor alone."

A main stumbling block in multisensor information fusion is the potential unknown correlations (or cross covariance) of the measurements at different sensors. This prevents the use of Kalman-filter-based fusion techniques, which need to know the joint distribution of the sensor measurement noises. It is also known that naive fusion of the correlated measurements by ignoring their correlations may lead to vastly incorrect estimation results [19]. To deal with the unknown correlation problem, the so-called covariance intersection (CI) algorithm became available. This algorithm was first proposed by Uhlmann in 1996 [2] and Julier and Uhlmann in 1997 [3]. Since its invention, a wide range of applications has been found, as cited in [4]–[9], [11]–[18]. Also, many modifications and theoretical interpretations of the CI method have been provided since its invention, including [5], [10], [15], [20], [22]–[26]. In particular, it is shown in [10] that CI is the optimal bounding algorithm in certain sense for two estimates under completely unknown correlations. See [18] and [26] for a larger list of references and their descriptions.

The CI algorithm plays a specially important role in distributed estimation and distributed fusion applications, where, instead of gathering the measurements of all the sensors and processing them centrally, each local sensor (or node) processes its own data and exchanges the local estimate with their neighboring nodes sensors so that global estimates are fused using these local estimates, see, e.g., [7], [13], [14], [27]–[29]. Distributed estimation and distributed fusion have less communicational burden, higher survivability, better flexibility, and reliability [26]. More importantly, these distributed algorithms tend to be more scalable to large networks.

The standard CI algorithm deals with the problem of mixing two random vectors (RVs) in an unbiased linear manner and gives an upper bound on the covariance of the mixed RV. This algorithm has two main drawbacks.

1) The algorithm gives an upper bound on the covariance of the sum of two correlated RVs. Many interpretations of the algorithm have been made to justify it. But the conservatism of the upper bound has not been clearly characterized. It is not clear either under what conditions this upper bound is tight.

Manuscript received February 14, 2017; revised May 23, 2017; accepted June 12, 2017. Date of publication June 21, 2017; date of current version February 26, 2018. This work was supported by the National Natural Science Foundation of China under Grant 61271210 and Grant 61633014. Recommended by Associate Editor Z. Chen. *(Corresponding author: Qianqian Cai.)*

Z. Wu and Q. Cai are with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China, and also with the Guangdong Key Laboratory of IoT Information Technology, Guangzhou 510006, China (e-mail: zzwu@scut.edu.cn; qianqian.cai@outlook.com).

M. Fu is with the School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW 2308, Australia, and also with the Guangdong Key Laboratory of IoT Information Technology, Guangzhou 510006, China (e-mail: minyue.fu@newcastle.edu.au).

Digital Object Identifier 10.1109/TAC.2017.2718243

2) There is no restriction on the degree of correlation, meaning that the result does not apply to *partially correlated* RVs.

Partial correlations are important to deal with because the very reason for fusion of multiple estimates is that they contain some uncorrelated information, thus treating them with "maximal" correlations is certainly not an ideal approach. One example of partial correlation is that two sensors may measure the state of the same object, but they contain independent measurement noises, which would induce a partial correlation between their estimates. Another example is when two estimates are such that each contains an independent component and a common component, which is the case motivating the so-called split CI algorithm, see [20]. The third example is the well-celebrated distributed statistical learning algorithm (an excellent classical example of data fusing algorithms) called *belief propagation* (BP) when applied to a networked system with a cyclic graph [30], [31]. In a BP algorithm, local estimates and their covariances at each subsystem are constantly shared and mixed with those from the neighboring subsystems, and feedback of the mixed information through loops would cause partial correlations. Partial correlations caused by feedback of estimates are also common in distributed Kalman-filter-based fusion algorithms [7] and distributed least-squares estimation [32].

Generalizations of the CI algorithm exist to handle partial correlations. For example, the split CI algorithm mentioned above was proposed in [20] and studied in more details in [15] and [21] to allow the covariances of each RV to be composed of a common component. This applies to the case where the error in the estimates can be decomposed into two mutually independent components. Unfortunately, many partially correlated estimates do not have such a nice decomposition. Hanebeck *et al.* [33], and Reece and Roberts [34] consider the joint covariance of two RVs with a given correlation coefficient bound and provide an upper bound for the joint covariance to improve the CI solution, but its optimality property is not studied.

This paper aims to deal with the aforementioned drawbacks. We first provide a tight result on CI for two RVs. This result gives an insight into the conservatism of the standard CI algorithm. On the other hand, we show that the standard CI algorithm is indeed optimal when a covariance upper bound is needed for the mixed RV *and* the trace of this upper bound is minimized. We then introduce the *Pearson's correlation coefficient bound* to describe the degree of correlation permitted by RVs and consider the CI problem for two RVs with a given correlation coefficient bound. The optimal solution is given when a covariance upper bound is needed for the mixed RV and the trace of this upper bound is minimized. This result generalizes the standard CI algorithm to two partially correlated RVs and gives an exact characterization of its optimality property. We also provide an alternating projection algorithm to solve the optimal scaling parameter and mixing gain matrices.

Another major contribution of this paper is to generalize the results to multiple RVs. We first introduce an appropriate generalization of the correlation coefficient bound to multiple RVs. We then provide a CI upper bound for the mixed RV for a given correlation coefficient bound. This CI upper bound resembles the known CI upper bound for the multiple RV case but involves more complicated parameterization to handle the partial correlations. In deriving this result, we also point out the fundamental technical difficulty in obtaining a tight upper bound for the multiple RV case. To minimize the proposed CI upper bound, we also study the optimal mixing gain matrices for the RVs and the optimal scaling parameters for the covariances. The convexity property of the proposed CI upper bound is presented, which is then used to derive an alternating projection algorithm to solve the optimal scaling parameters and mixing gain matrices for the multiple RV case.

The rest of the paper is organized as follows: Section II provides a tight bound on CI with the purpose to reveal exactly the conservatism of the standard CI algorithm. Section III solves the problem of CI for two partially correlated RVs. Section IV generalizes the results in Section III to multiple RVs. Three illustrating examples are given, two in Section II and one in Section IV. Section V concludes the paper.

II. Tighter Trace Bound on the Covariance of Fused Estimates

To motivation this section, we first revisit the problem formulation and solution for CI. Let a and b be two unbiased estimates of an RV x. The autocovariances $P_a = \mathbb{E}\{aa^T\} > 0$ and $P_b = \mathbb{E}\{bb^T\} > 0$ are known but their cross correlation $P_{ab} = \mathbb{E}\{ab^T\}$ is unknown. The fusion of the two estimates above involves constructing a linearly combined estimate

$$c = K_1 a + K_2 b \tag{1}$$

with some constant matrices K_1 and K_2 constrained by $K_1 + K_2 = I$ to ensure that c is unbiased. The problem of CI is twofold [5].

1) For a given pair of K_1 and K_2, find a covariance upper bound $\hat{P} \geq P_c = \mathbb{E}\{cc^T\}$.
2) Find the pair of K_1 and K_2, such that the upper bound $\hat{P}$ is optimal in some sense, e.g., minimal trace or determinant (we will focus on the trace in this paper).

The standard solution is given by the CI algorithm [3]:

$$\hat{P}^{-1} = \omega P_a^{-1} + (1-\omega) P_b^{-1}, \tag{2}$$

$$K_1 = \omega \hat{P} P_a^{-1}, \quad K_2 = (1-\omega)\hat{P} P_b^{-1} \tag{3}$$

with $\omega \in [0,\ 1]$ which can be easily searched to minimize the trace or determinant of $\hat{P}$.

The motivation for this section stems from the fact that such "optimized" K_1 and K_2 do not necessarily produce a fused estimate c in (1) with a minimal P_c, e.g., in terms of its trace. The reason for this observation is that K_1 and K_2 are actually optimized for the upper bound $\hat{P}$, not directly for the covariance of P_c.

The purpose of this section is to give a tighter bound on $\mathrm{Tr}(P_c)$ than offered by CI. To this end, we will first assume that K_1 and K_2 are given, and for notational convenience, let K_1 and K_2 be absorbed (or by viewing $K_1 a$ as a and $K_2 b$ as b). We will give a tight trace bound on P_c for $c = a + b$, i.e., we will compute the exact worst-case trace of the covariance for the sum of these two RVs. This result will then be extended to the case where the two correlated RVs are mixed using linear combination to give a tighter bound than offered by the standard CI

algorithm. Our analysis is based on the so-called *nuclear norm* (or *trace norm*) for matrices [35].

Definition 1: For a matrix $A \in \mathbb{R}^{m\times n}$, denote by $\|A\|_\star$ its *nuclear norm* (also known as *trace norm*) which is defined by

$$\|A\|_\star = \sum_{i=1}^{\min\{m,n\}} \sigma_i(A) = \mathrm{Tr}\left(\sqrt{A^TA}\right) \tag{4}$$

where $\sigma_i(A)$ are the singular values of A and $\mathrm{Tr}(\cdot)$ is the trace operator.

The nuclear norm enjoys the following basic properties. These properties are well known. For example, Property 3 is the basis of semidefinite programming formulations of the trace norm (see, e.g., [35]), and Property 4 can be found in [36]. But we provide a proof for completeness.

Lemma 1: The properties below hold for the nuclear norm:

1) $\|A\|_\star = \|A^T\|_\star$ for any matrix A.
2) $\|A\|_\star = \mathrm{Tr}(A)$ for any symmetric and positive semidefinite A.
3) $2\|AB^T\|_\star \le \mathrm{Tr}(A^TA) + \mathrm{Tr}(B^TB)$ for any dimension-compatible matrices A and B.
4) $\|AB^T\|_\star \le \sqrt{\mathrm{Tr}(A^TA)\mathrm{Tr}(B^TB)}$ for any dimension-compatible matrices A and B.

Proof: The first two properties follow directly from the definition. The inequality above is verified as follows: Take the singular value decomposition $AB^T = U\Sigma V^T$ with unitary matrices U and V and diagonal $\Sigma > 0$. Then

$$\|AB^T\|_\star = \mathrm{Tr}(\Sigma) = \mathrm{Tr}(U^TAB^TV) = \mathrm{Tr}(AB^T\Gamma)$$

with $\Gamma = VU^T$. Note that Γ is unitary. It follows from the first property that

$$\begin{aligned} 2\|AB^T\|_\star &= \mathrm{Tr}(AB^T\Gamma + \Gamma^TBA^T) \\ &\le \mathrm{Tr}(A^TA + B^T\Gamma\Gamma^TB) \\ &= \mathrm{Tr}(A^TA + B^TB). \end{aligned}$$

Finally, using the third property but taking $\tilde{A} = \gamma^{1/2}A$ and $\tilde{B} = \gamma^{-1/2}B$ with some scalar $\gamma > 0$ (to be specified), we obtain

$$2\|AB^T\|_\star \le \gamma\mathrm{Tr}(A^TA) + \gamma^{-1}\mathrm{Tr}(B^TB).$$

If $\mathrm{Tr}(A^TA) = 0$, we have $2\|AB^T\|_\star \le \gamma^{-1}\mathrm{Tr}(B^TB)$ for any γ, thus $2\|AB^T\|_\star = 0 = \sqrt{\mathrm{Tr}(A^TA)\mathrm{Tr}(B^TB)}$. If $\mathrm{Tr}(A^TA) > 0$, taking

$$\gamma = \left(\mathrm{Tr}(A^TA)\right)^{-1/2}\left(\mathrm{Tr}(B^TB)\right)^{1/2}$$

the above becomes

$$2\|AB^T\|_\star \le 2\sqrt{\mathrm{Tr}(A^TA)\mathrm{Tr}(B^TB)}.$$

This completes the proof. ■

The following result gives a tight bound on the covariance of the sum of two RVs with a unknown cross correlation.

Lemma 2: Given two correlated zero-mean RVs $a, b \in \mathbb{R}^n$, suppose $\mathbf{E}\{aa^T\} = P_a > 0$ and $\mathbf{E}\{bb^T\} = P_b > 0$ are known but $\mathbf{E}\{ab^T\} = P_{ab}$ is unknown. Let $P_a = XX^T$ and $P_b = YY^T$ be any decompositions of P_a and P_b, respectively, with $X, Y \in \mathbb{R}^{n\times n}$. Then

$$\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{(a+b)(a+b)^T\}) = \|P_a\|_\star + \|P_b\|_\star + 2\|Y^TX\|_\star.$$

Proof: For the given a, b, we have

$$\mathrm{Tr}(\mathbf{E}\{(a+b)(a+b)^T\}) = \mathrm{Tr}(P_a + P_b + P_{ab} + P_{ab}^T).$$

Since P_a and P_b are symmetric, we have

$$\mathrm{Tr}(P_a + P_b) = \mathrm{Tr}(P_a) + \mathrm{Tr}(P_b) = \|P_a\|_\star + \|P_b\|_\star.$$

Also note that $\mathrm{Tr}(P_{ab}^T) = \mathrm{Tr}(P_{ab})$. Therefore, it suffices to show that

$$\max_{P_{ab}} \mathrm{Tr}(P_{ab}) = \|Y^TX\|_\star.$$

Note that P_{ab} is naturally constrained by

$$\begin{bmatrix} P_a & P_{ab} \\ P_{ab}^T & P_b \end{bmatrix} \ge 0$$

which is the same as $P_{ab}P_b^{-1}P_{ab}^T \le P_a$. Using the decompositions of P_a and P_b, the above becomes

$$(X^{-1}P_{ab}(Y^T)^{-1})(X^{-1}P_{ab}(Y^T)^{-1})^T \le I.$$

This implies that

$$X^{-1}P_{ab}(Y^T)^{-1} = \Gamma,\ \ \Gamma\Gamma^T \le I.$$

That is, P_{ab} is fully parameterized by

$$P_{ab} = X\Gamma Y^T,\ \ \Gamma\Gamma^T \le I.$$

Hence

$$\mathrm{Tr}(P_{ab}) = \mathrm{Tr}(X\Gamma Y^T) = \mathrm{Tr}((Y^TX)\Gamma).$$

Denote by $Y^TX = U\Sigma V^T$ the singular value decomposition of Y^TX, i.e., U and V are unitary matrices and $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ containing the singular values of Y^TX. Then

$$\mathrm{Tr}((Y^TX)\Gamma) = \mathrm{Tr}(U\Sigma V^T\Gamma) = \mathrm{Tr}(\Sigma V^T\Gamma U) = \mathrm{Tr}(\Sigma\Pi)$$

with $\Pi = V^T\Gamma U$. It is clear that $\Gamma\Gamma^T \le I$ if and only if $\Pi\Pi^T \le I$. Note that the diagonal elements π_{ii} of Π satisfies the constrains that $|\pi_{ii}| \le 1$, which follows from the simple calculation that

$$1 = e_i^Te_i \ge e_i^T\Pi\Pi^Te_i = \sum_{j=1}^n \pi_{ij}^2 \ge \pi_{ii}^2.$$

In the above, e_i is the column vector with all entries equal to 0 except that the ith entry is 1. Using the above

$$\mathrm{Tr}(\Sigma\Pi) = \sum_{i=1}^n \sigma_i\pi_{ii} \le \sum_{i=1}^n \sigma_i.$$

It is clear that the inequality is reached by choosing $\Pi = I$. Hence

$$\max_{P_{ab}} \mathrm{Tr}(P_{ab}) = \|Y^TX\|_\star.$$

This completes the proof. ■

Remark 1: Lemma 2 provides the solution for the maximum $\mathrm{Tr}(\mathbf{E}\{(a+b)(a+b)^T\})$ without giving its maximizer P_{ab} explicitly. It is actually known that the maximizer is the geometric mean of P_a and P_b, see [37] and references therein.

Remark 2: The standard CI algorithm is obtained based on the following upper bound (see [3]):

$$\max_{P_{ab}} \mathrm{Tr}(P_a + P_b + P_{ab} + P_{ab}^T) \leq (\sqrt{\mathrm{Tr}(P_a)} + \sqrt{\mathrm{Tr}(P_b)})^2.$$

To show the conservatism of this upper bound, we take the following example.

Example 1: Consider the case with

$$P_a = \mathrm{diag}\{1,\ 4\},\ \ P_b = \mathrm{diag}\{4,\ 1\}. \tag{5}$$

It is easy to compute that $(\sqrt{\mathrm{Tr}(P_a)} + \sqrt{\mathrm{Tr}(P_b)})^2 = 20$, whereas $\|P_a\|_\star + \|P_b\|_\star + 2\|Y^T X\|_\star = 18$.

The following is our first main result on the tight bound for the trace of the covariance for linearly fused estimates.

Theorem 1: Given two correlated zero-mean RVs $a, b \in \mathbb{R}^n$ with known $\mathbf{E}\{aa^T\} = P_a > 0$ and $\mathbf{E}\{bb^T\} = P_b > 0$ but unknown $\mathbf{E}\{ab^T\} = P_{ab}$, consider the following RV:

$$c = K_1 a + K_2 b$$

where $K_1, K_2 \in \mathbb{R}^{n\times n}$ are nonsingular matrices. Let $P_a = XX^T$ and $P_b = YY^T$ be any decompositions of P_a and P_b, respectively, with $X, Y \in \mathbb{R}^{n\times n}$. Then

$$\begin{aligned}\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{cc^T\}) = \ & \|K_1 P_a K_1^T\|_\star + \|K_2 P_b K_2^T\|_\star \\ & + 2\|Y^T K_2^T K_1 X\|_\star. \end{aligned} \tag{6}$$

Moreover, the following inequality holds:

$$\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{cc^T\}) \leq \left(\sqrt{\mathrm{Tr}(K_1 P_a K_1^T)} + \sqrt{\mathrm{Tr}(K_2 P_b K_2^T)}\right)^2.$$

Proof: Taking $\tilde{a} = K_1 a$ and $\tilde{b} = K_2 b$, we have $\mathbf{E}\{\tilde{a}\tilde{a}^T\} = P_{\tilde{a}} = K_1 P_a K_1^T$, $\mathbf{E}\{\tilde{b}\tilde{b}^T\} = P_{\tilde{b}} = K_2 P_b K_2^T$, $\mathbf{E}\{\tilde{a}\tilde{b}^T\} = P_{\tilde{a}\tilde{b}} K_1 P_{ab} K_2^T$. The first result is obtained by directly applying Lemma 2. The inequality follows by applying the fourth property of the nuclear norm in Lemma 1. More precisely, taking the matrices A and B in Lemma 1 as $A = Y^T K_2^T$ and $B = X^T K_1^T$, we have

$$\begin{aligned} & \|K_1 P_a K_1^T\|_\star + \|K_2 P_b K_2^T\|_\star + 2\|Y^T K_2^T K_1 X\|_\star \\ & \leq \left(\sqrt{\mathrm{Tr}(K_1 P_a K_1^T)} + \sqrt{\mathrm{Tr}(K_2 P_b K_2^T)}\right)^2. \end{aligned}$$

This completes the proof. ■

Example 2: We now apply Theorem 1 to show a tighter trace bound can be obtained than that given by the standard CI algorithm. Consider the two estimates a and b with autocovariances in (5). If we apply the standard CI algorithm, we obtain, from (2), that

$$\hat{P} = \mathrm{diag}\left\{\frac{1}{\omega + 0.25(1-\omega)},\ \frac{1}{0.25\omega + (1-\omega)}\right\}.$$

By minimizing $\mathrm{Tr}(\hat{P})$, we get the optimal $\omega = 0.5$, and the corresponding $\mathrm{Tr}(\hat{P}) = 3.2$ and

$$K_1 = \mathrm{diag}\{0.8,\ 0.2\};\ \ K_2 = \mathrm{diag}\{0.2,\ 0.8\}. \tag{7}$$

Also interestingly, if we take the above K_1 and K_2 and apply (5), we actually obtain

$$\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{cc^T\}) = 2.88.$$

In comparison, we apply Theorem 1 and for simplicity, we restrict $K_1 = \mathrm{diag}\{k_1,\ k_2\}$ and $K_2 = \mathrm{diag}\{1-k_1,\ 1-k_2\}$. Note that we can take $X = \mathrm{diag}\{1,\ 2\}$ and $Y = \mathrm{diag}\{2,\ 1\}$. Applying (6) and (4) gives

$$\begin{aligned}\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{cc^T\}) &= k_1^2 + 4k_2^2 + 4(1-k_1)^2 + (1-k_2)^2 \\ &\quad + 4(|k_1(1-k_1)| + |k_2(1-k_2)|) \\ &= (|k_1| + 2|1-k_1|)^2 + (2|k_2| + |1-k_2|)^2.\end{aligned}$$

Minimizing the above gives the optimal $k_1 = 1$ and $k_2 = 0$, which yields

$$\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{cc^T\}) = 2;\ \ K_1 = \mathrm{diag}\{1,\ 0\};\ \ K_2 = \mathrm{diag}\{0,\ 1\}.$$

It is clear that this is a much tighter bound than that given by the standard CI algorithm. It can be verified (although not shown here) that this is also the optimal solution for general (not necessarily diagonal) K_1 and K_2.

III. CI With Partial Correlations

In this section, we address the second shortcoming of the standard CI algorithm by generalizing it to RVs with partial correlations. This allows less conservative covariance bounds to be obtained. We also generalize the tight bound result (see Theorem 1) to RVs with partial correlations.

Our first task is to characterize the "degree" of correlation between two RVs. To this end, we introduce the so-called *Pearson's correlation coefficient*.

A. Pearson's Correlation Coefficient

For two random variables (scalars) a and b with means μ_a and μ_b and variances σ_a^2 and σ_b^2, respectively, their correlation can be characterized by the well-known Pearson's correlation coefficient defined by

$$\rho_{ab} = \frac{\mathbf{E}\{(a-\mu_a)(b-\mu_b)\}}{\sigma_a \sigma_b}.$$

Generalizing this notion to RVs, we let $a, b \in \mathbb{R}^n$ with means μ_a and μ_b and covariances $P_a > 0$ and $P_b > 0$, respectively. Let $P_a = XX^T$ and $P_b = YY^T$ be any decompositions with $X, Y \in \mathbb{R}^{n\times n}$. Then, the *correlation coefficient bound* for a and b is given by

$$\rho_{ab} = \sigma_{\max}\left(X^{-1}\mathbf{E}\{(a-\mu_a)(b-\mu_b)^T\}(Y^T)^{-1}\right) \tag{8}$$

where $\sigma_{\max}$ denotes the maximum singular value. Although X and Y are not unique, it is clear that ρ_{ab} is the same for any X, Y. We note that this notion of correlation coefficient bound is also used in [33] and [34].

The definition of correlation coefficient bound in (8) has a simple interpretation, as stated in the result below.

Lemma 3: Given a scalar constant $0 \leq \rho \leq 1$ and two RVs $a, b \in \mathbb{R}^n$ with means μ_a and μ_b and covariances $P_a > 0$ and

$P_b > 0$, respectively, their correlation coefficient ρ_{ab}, as defined in (8), has $\rho_{ab} \leq \rho$ if and only if

$$\begin{bmatrix} \rho P_a & P_{ab} \\ P_{ab}^T & \rho P_b \end{bmatrix} \geq 0. \tag{9}$$

Proof: From (8), it is clear that if $\rho = 0$, then $\rho_{ab} = 0$ which implies that $P_{ab} = 0$, hence (9) holds. Conversely, if (9) holds for $\rho = 0$, then $P_{ab} = 0$ which implies $\rho_{ab} = 0$ and $\rho = 0$. Therefore, we only need to consider $\rho > 0$. For this case, $\rho_{ab} \leq \rho$ if and only if

$$\sigma_{\max}(P_a^{-1/2} P_{ab} P_b^{-1/2}) \leq \rho$$

which is equivalent to

$$P_a^{-1/2} P_{ab} P_b^{-1} P_{ab}^T P_a^{-1/2} \leq \rho^2 I$$

or

$$P_{ab}(\rho P_b)^{-1} P_{ab}^T \leq \rho P_a.$$

Using Schur complement, the above is equivalent to (9). ■

The result above means that ρ_{ab} is the minimum value of ρ to satisfy (9). For this reason, we will call any $\rho \geq \rho_{ab}$ a *correlation coefficient upper bound*. In practice, it is often difficult to obtain the exact ρ_{ab}, but obtaining an upper bound may often be easier. In the sequel, we will assume that an correlation coefficient upper bound is available.

We now generalize Lemma 2.

Lemma 4: Given two correlated zero-mean RVs $a, b \in \mathbb{R}^n$, suppose $\mathbf{E}\{aa^T\} = P_a > 0$ and $\mathbf{E}\{bb^T\} = P_b > 0$ are known and their correlation coefficient is bounded by $\rho \in [0,\ 1]$. Let $P_a = XX^T$ and $P_b = YY^T$ be any decompositions of P_a and P_b, respectively, with $X, Y \in \mathbb{R}^{n\times n}$. Then

$$\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{(a+b)(a+b)^T\}) = \|P_a\|_\star + \|P_b\|_\star + 2\rho\|Y^T X\|_\star.$$

Proof: The proof is modified from that of Lemma 2. As in that proof, it suffices to show that

$$\max_{P_{ab}} \mathrm{Tr}(P_{ab}) = \rho\|Y^T X\|_\star. \tag{10}$$

This is done by invoking (9), which, by Schur complement, is the same as $P_{ab} P_b^{-1} P_{ab}^T \leq \rho^2 P_a$. Following the same decomposition for P_a and P_b as in that proof, we get the same characterization for P_{ab} as $P_{ab} = X\Gamma Y^T$, except the new constraint $\Gamma\Gamma^T \leq \rho^2 I$. Taking the singular value decomposition of $Y^T X = U\Sigma V^T$ and the defintion of $\Pi = V^T \Gamma U$, the new constraint on Π becomes $\Pi\Pi^T \leq \rho^2 I$. Following the same steps as in that proof, we also obtain

$$\mathrm{Tr}(P_{ab}) = \mathrm{Tr}((Y^T X)\Gamma) = \mathrm{Tr}(\Sigma\Pi) \leq \rho\mathrm{Tr}(\Sigma)$$

with the equality reached when $\Pi = \rho I$. Hence, (10) holds. ■

The result above leads to a direct generalization of Theorem 1 to correlated RVs with a given correlation coefficient bound. The proof is omitted.

Theorem 2: Given two correlated zero-mean RVs $a, b \in \mathbb{R}^n$ with known $\mathbf{E}\{aa^T\} = P_a > 0$ and $\mathbf{E}\{bb^T\} = P_b > 0$ and their correlation coefficient is bounded by $\rho \in [0,\ 1]$, consider the following RV

$$c = K_1 a + K_2 b$$

where $K_1, K_2 \in \mathbb{R}^{n\times n}$ are nonsingular matrices. Let $P_a = XX^T$ and $P_b = YY^T$ be any decompositions of P_a and P_b, respectively, with $X, Y \in \mathbb{R}^{n\times n}$. Then

$$\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{cc^T\}) = \|K_1 P_a K_1^T\|_\star + \|K_2 P_b K_2^T\|_\star + 2\rho\|Y^T K_2^T K_1 X\|_\star. \tag{11}$$

Moreover, the following inequality holds:

$$\max_{P_{ab}} \mathrm{Tr}(\mathbf{E}\{cc^T\}) \leq \left(\sqrt{\mathrm{Tr}(K_1 P_a K_1^T)} + \sqrt{\mathrm{Tr}(K_2 P_b K_2^T)}\right)^2 - (1-\rho)\|Y^T K_2^T K_1 X\|_\star.$$

Theorem 2 gives a way to optimize the weighting matrices K_1 and K_2. That is, they can be computed by minimizing the right-hand side of (11). However, simple optimization algorithms are difficult to obtain due to the presence of the trace norm. Next, we give a generalized CI algorithm to handle the partially correlated estimates. For this, we need to introduce a technical lemma first.

Lemma 5: Given two symmetric and positive-definite covariance matrices $P_a, P_b \in \mathbb{R}^{n\times n}$ and a correlation coefficient bound $\rho \in [0,\ 1]$, consider the set of zero-mean RV pairs

$$\Omega = \{(a,b) : \mathbf{E}\{aa^T\} = P_a, \mathbf{E}\{bb^T\} = P_b, \rho_{ab} \leq \rho\}.$$

Let $K_1, K_2 \in \mathbb{R}^{n\times n}$ be given nonsingular matrices and consider the RV

$$c = K_1 a + K_2 b \tag{12}$$

for any $(a,b) \in \Omega$ and denote $P_c = \mathbf{E}\{cc^T\}$. Then, the optimal solution to the following problem

$$\min \mathrm{Tr}(\hat{P}) \text{ subject to } \hat{P} \geq P_c,\ \ \forall (a,b) \in \Omega$$

is given by

$$\hat{P} = (1+\gamma\rho) K_1 P_a K_1^T + (1+\gamma^{-1}\rho) K_2 P_b K_2^T \tag{13}$$

with

$$\gamma = \left(\mathrm{Tr}(K_1 P_a K_1^T)\right)^{-1/2} \left(\mathrm{Tr}(K_2 P_b K_2^T)\right)^{1/2} \tag{14}$$

and the corresponding minimum $\mathrm{Tr}(\hat{P})$ is given by

$$\mathrm{Tr}(\hat{P}) = \mathrm{Tr}(K_1 P_a K_1^T) + \mathrm{Tr}(K_2 P_b K_2^T) + 2\rho\left(\mathrm{Tr}(K_1 P_a K_1^T)\mathrm{Tr}(K_2 P_b K_2^T)\right)^{1/2}. \tag{15}$$

Proof: From the definition of c, we have

$$\begin{aligned} P_c &= K_1 P_a K_1^T + K_2 P_b K_2^T + K_1 P_{ab} K_2^T + K_2 P_{ab}^T K_1^T \\ &= K_1 P_a K_1^T + K_2 P_b K_2^T \\ &\quad + K_1 P_a^{1/2} \Gamma P_b^{1/2} K_2^T + K_2 P_b^{1/2} \Gamma^T P_a^{1/2} K_1^T \end{aligned}$$

where $\Gamma = P_a^{-1/2} P_{ab} P_b^{-1/2}$ with $\Gamma^T\Gamma \leq \rho^2 I$. Taking $A = (\rho\gamma)^{1/2} K_1 P_a^{1/2}$ and $B = (\rho\gamma)^{-1/2} \Gamma K_2 P_b^{1/2}$ for any $\gamma > 0$ and using $AB^T + BA^T \leq AA^T + BB^T$, it follows that

$$P_c \leq (1+\gamma\rho) K_1 P_a K_1^T + (1+\gamma^{-1}\rho) K_2 P_b K_2^T.$$

That is, the right-hand side above serves as an upper bound for P_c for any $\gamma > 0$. It is easy to verify that minimizing the trace of

the right-hand side above yields the solution to γ in (14), and the corresponding minimum trace is given by (15). Now we claim that $\hat{P}$ in (13) with γ in (14) is the upper bound of P_c with the minimum trace. Indeed, if this is not true, there will be another upper bound $\tilde{P}$ of P_c with its trace less than that given by (15). Then, using the well-known S-procedure [38] (also see [39]), $P_c \le \tilde{P}$ for all Γ with $\Gamma^T\Gamma \le \rho^2 I$ if and only if there exists some scalar $\gamma > 0$, such that

$$(1+\gamma\rho)K_1P_aK_1^T + (1+\gamma^{-1}\rho)K_2P_bK_2^T \le \tilde{P}.$$

Taking the trace on both sides above, note that the left-hand side is no less than that given by (15). It is clear that this will contradict the fact that the trace of $\tilde{P}$ is less than that given by (15). Hence, the upper bound $\hat{P}$ for P_c with the minimum trace is given by (13) with γ in (14). ■

We see from Lemma 5 that the covariance upper bound (13) given by the standard CI algorithm is tight when all the RV pairs $(a,\ b)$ in the set Ω are considered.

Using Lemma 5, we obtain the following generalized CI algorithm to handle partially correlated estimates.

Theorem 3: Given two symmetric and positive-definite covariance matrices $P_a, P_b \in \mathbb{R}^{n\times n}$ and a correlation coefficient bound $\rho \in [0,\ 1]$, consider the set of RV pairs

$$\Omega = \{(a,b) : \mathbf{E}\{a\} = \mathbf{E}\{b\}, \mathrm{cov}(a) = P_a, \\ \mathrm{cov}(b) = P_b, \rho_{ab} \le \rho\}. \quad (16)$$

For any $K_1, K_2 \in \mathbb{R}^{n\times n}$, consider the unbiased linear combination (12). Then, the optimal solution to the following problem

$$\min\ \ \mathrm{Tr}(\hat{P})\ \textit{subject to} \\ \hat{P} \ge \mathrm{cov}(c),\ \ \forall (a,b)\in\Omega;\ \ K_1+K_2 = I \quad (17)$$

is given by

$$\hat{P} = \left((1+\gamma\rho)^{-1}P_a^{-1} + (1+\gamma^{-1}\rho)^{-1}P_b^{-1}\right)^{-1} \quad (18)$$

and

$$K_1 = (1+\gamma\rho)^{-1}\hat{P}P_a^{-1};\ \ K_2 = (1+\gamma^{-1}\rho)^{-1}\hat{P}P_b^{-1} \quad (19)$$

with some $\gamma > 0$. Alternatively, the solution above is given by

$$\hat{P} = \left(w_1P_a^{-1} + w_2P_b^{-1}\right)^{-1}$$

and

$$K_1 = w_1\hat{P}P_a^{-1};\ \ K_2 = w_2\hat{P}P_b^{-1}$$

with

$$w_1^{-1} = \delta_1^{-1}\rho + 1 - \rho;\ \ w_2^{-1} = \delta_2^{-1}\rho + 1 - \rho$$

subject to $\delta_1 > 0, \delta_2 > 0$ and $\delta_1 + \delta_2 = 1$.

Proof: Following from Lemma 5 and its proof, for fixed K_1, K_2, the optimal upper bound for $\mathrm{cov}(c)$ is given by

$$P(\gamma) = (1+\gamma\rho)K_1P_aK_1^T + (1+\gamma^{-1}\rho)K_2P_bK_2^T \quad (20)$$

with γ chosen to minimize its trace. Therefore, the optimal K_1, K_2 is obtained by

$$\min_{K_1,K_2}\ \min_{\gamma>0} \mathrm{Tr}(P(\gamma))$$

subject to $K_1 + K_2 = I$. Since minimizations are done with respect to both (K_1, K_2) and γ, we can swap them the order of minimizations. Swapping the minimization steps and fixing γ, we first consider

$$\min_{K_1,K_2}\ \mathrm{Tr}(P(\gamma)) \text{ subject to } K_1 + K_2 = I.$$

Replacing $K_2 = I - K_1$ and using the fact that

$$\frac{\partial \mathrm{Tr}(XPX^T)}{\partial X} = 2XP$$

we have

$$\frac{\partial \mathrm{Tr}(P(\gamma))}{\partial K_1} = 2((1+\gamma\rho)K_1P_a + (1+\gamma^{-1}\rho)(K_1 - I)P_b).$$

Setting it to zero, we obtain the optimal solution for K_1 as

$$K_1 = (1+\gamma^{-1}\rho)P_b((1+\gamma\rho)P_a + (1+\gamma^{-1}\rho)P_b)^{-1} \\ = ((1+\gamma\rho)^{-1}P_a^{-1} + (1+\gamma^{-1}\rho)^{-1}P_b^{-1})^{-1} \\ \cdot (1+\gamma\rho)^{-1}P_a^{-1}.$$

It follows that

$$K_2 = ((1+\gamma\rho)^{-1}P_a^{-1} + (1+\gamma^{-1}\rho)^{-1}P_b^{-1})^{-1} \\ \cdot (1+\gamma^{-1}\rho)^{-1}P_b^{-1}.$$

Plugging in K_1 and K_2 above into $P(\gamma)$ yields

$$P(\gamma) = ((1+\gamma\rho)^{-1}P_a^{-1} + (1+\gamma^{-1}\rho)^{-1}P_b^{-1})^{-1}.$$

The alternative solution is given by taking $w_1 = (1+\gamma\rho)^{-1}$, $w_2 = (1+\gamma^{-1}\rho)^{-1}$, $\delta_1 = (1+\gamma)^{-1}$ and $\delta_2 = (1+\gamma^{-1})^{-1}$. It is straightforward to verify that

$$w_1^{-1} = \delta_1^{-1}\rho + 1 - \rho;\ \ w_2^{-1} = \delta_2^{-1}\rho + 1 - \rho$$

with $\delta_1 > 0, \delta_2 > 0$ and $\delta_1 + \delta_2 = 1$ for $\gamma > 0$. ■

Remark 3: When $\rho = 0$ (i.e., a and b are uncorrelated), the solution above is optimized by taking any γ, which gives $\hat{P} = (P_a^{-1} + P_b^{-1})^{-1}$. When $\rho = 1$ (i.e., a and b are maximally correlated), the solution above becomes the standard CI algorithm, and our result coincides with a recent result (Theorem 3 in [10]) about the optimality of the CI algorithm. For a general $0 < \rho < 1$, the solution $\hat{P}$ in (18) decreases as ρ decreases because both $1/(1+\gamma\rho)$ and $1/(1+\gamma^{-1}\rho)$ monotonically increase as ρ decreases.

Our next result addresses the problem of searching for the optimal γ to minimize the trace of $\hat{P}$ in (18).

Lemma 6: Given any symmetric and positive-definite matrices P_a and P_b and correlation coefficient bound $0 < \rho \le 1$, consider the matrix function $\hat{P}$ of γ as defined in (18). Suppose $\gamma \in (0,\infty)$ is a stationary point of $\mathrm{Tr}(\hat{P})$, then it must be the unique global minimum point in $(0,\ \infty)$.

Proof: Denote

$$\Delta = \hat{P}^{-1} = (1+\gamma\rho)^{-1}P_a^{-1} + (1+\gamma^{-1}\rho)^{-1}P_b^{-1}.$$

Let γ be a given stationary point of $\mathrm{Tr}(\hat{P})$. Using the fact that

$$\frac{\partial \Delta^{-1}}{\partial\gamma} = -\Delta^{-1}\frac{\partial\Delta}{\partial\gamma}\Delta^{-1}$$

we have

$$\frac{\partial}{\partial\gamma}\mathrm{Tr}(\Delta^{-1}) = -\mathrm{Tr}\left(\Delta^{-1}\frac{\partial\Delta}{\partial\gamma}\Delta^{-1}\right) = 0;$$

$$\frac{\partial^2}{\partial\gamma^2}\mathrm{Tr}(\Delta^{-1}) = 2\mathrm{Tr}\left(\Delta^{-1}\frac{\partial\Delta}{\partial\gamma}\Delta^{-1}\frac{\partial\Delta}{\partial\gamma}\Delta^{-1}\right) - \mathrm{Tr}\left(\Delta^{-1}\frac{\partial^2\Delta}{\partial\gamma^2}\Delta^{-1}\right).$$

Note that

$$\frac{\partial\Delta}{\partial\gamma} = -\frac{\rho}{(1+\gamma\rho)^2}P_a^{-1} + \frac{\rho}{(\gamma+\rho)^2}P_b^{-1};$$

$$\frac{\partial^2\Delta}{\partial\gamma^2} = \frac{2\rho^2}{(1+\gamma\rho)^3}P_a^{-1} - \frac{2\rho}{(\gamma+\rho)^3}P_b^{-1} = \frac{2\rho^2}{(1+\gamma\rho)^3}P_a^{-1} - \frac{2}{\gamma+\rho}\left(\frac{\partial\Delta}{\partial\gamma} + \frac{\rho}{(1+\gamma\rho)^2}P_a^{-1}\right) = -\frac{2\rho(1-\rho^2)}{(1+\gamma\rho)^3(\gamma+\rho)}P_a^{-1} - \frac{2}{\gamma+\rho}\frac{\partial\Delta}{\partial\gamma}.$$

It follows that

$$\frac{\partial^2}{\partial\gamma^2}\mathrm{Tr}(\Delta^{-1}) = 2\mathrm{Tr}\left(\Delta^{-1}\frac{\partial\Delta}{\partial\gamma}\Delta^{-1}\frac{\partial\Delta}{\partial\gamma}\Delta^{-1}\right) + \frac{2}{\gamma+\rho}\mathrm{Tr}\left(\Delta^{-1}\frac{\partial\Delta}{\partial\gamma}\Delta^{-1}\right) + \frac{2\rho(1-\rho^2)}{(1+\gamma\rho)^3(\gamma+\rho)}\mathrm{Tr}\left(\Delta^{-1}P_a^{-1}\Delta^{-1}\right).$$

Note in the above that the second term is zero and the first and third terms are nonnegative. The third term is zero only if $\rho = 1$. Also, the first term is zero only if

$$\frac{\partial}{\partial\gamma}\Delta = 0$$

which holds only when $P_b = \alpha P_a$ for some scalar $\alpha > 0$. Thus, the sum above is zero only if $\rho = 1$ and $P_b = \alpha P_a$ for some scalar $\alpha > 0$. In this case

$$\hat{P} = \left(\frac{1}{1+\gamma} + \frac{1}{\alpha(1+\gamma^{-1})}\right)^{-1}P_a = \frac{1+\gamma}{1+\alpha^{-1}\gamma}P_a.$$

It can be easily verified that the term $(1+\gamma)(1+\alpha^{-1}\gamma)^{-1}$ does not have a stationary point in $(0,\ \infty)$. Hence, at the given stationary point $\gamma \in (0,\ \infty)$, the second-order derivative of $\mathrm{Tr}(\hat{P})$ must be positive. That is, γ must be a strict local minimum point. It is clear that this is also a global minimum point and a unique one because otherwise there will be at least two distinct minimum points, which implies that there will be another local maximum point (an equilibrium point) in between which is not possible by what we have just shown. ■

Remark 4: We have excluded $\rho = 0$ in the above lemma. For $\rho = 0$, γ does not play a role and the resulting $\hat{P} = (P_a^{-1} + P_b^{-1})^{-1}$ as mentioned in Remark 3.

Following the lemma above, we give an alternating projection algorithm for optimizing the γ.

The convergence property of the algorithm above is stated below.

Algorithm 1: (Alternating Projection for Optimizing γ).

Step 1: Initialize $\gamma = 1$;
Step 2: Compute $\hat{P}$, K_1 and K_2 as in (18)–(19);
Step 3: Compute γ as in (14) and return to Step 2 until convergence.

Theorem 4: Given any symmetric and positive-definite matrices P_a and P_b and correlation coefficient bound $0 < \rho \leq 1$, the algorithm above convergences to the optimal solution for minimizing the trace of $\hat{P}$ in (18).

Proof: Consider the upper bound $P(\gamma)$ as defined in (20). It is clear from Theorem 3 that for a given $\gamma > 0$, (18) and (19) give the optimal solution for K_1 and K_2 to minimize $\mathrm{Tr}(P(\gamma))$. Then, from Theorem 1, for the given K_1 and K_2, $P(\gamma) = \hat{P}$ is minimized by taking (14). Each iteration (Steps 2, 3) above will reduce $\mathrm{Tr}(\hat{P})$ unless γ is an interior stationary point, which by Lemma 4, is the global minimum point. Hence, γ converges to either an interior stationary point or a boundary point (0 or ∞). In either case, $\mathrm{Tr}(\hat{P})$ will be minimized. ■

IV. Generalization to Multiple RVs

In this section, we generalize the results in the previous section to multiple RVs. We first provide a reinterpretation of Lemma 5 and Theorem 3, which will allow generalization to multiple RVs. The key is to exploit Lemma 3, which leads to a natural generalization of the notion of correlation coefficient bound to multiple RVs.

Definition 2: Given a constant vector $\rho = [\rho_1\ \rho_2\ \ldots\ \rho_m]$ with $0 \leq \rho_i \leq 1$ for all i and RVs $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$, the set of these RVs is said to have correlation coefficient bound of ρ if

$$\mathbf{P}_m^{(\rho)} = \begin{bmatrix} \rho_1 P_1 & P_{12} & P_{13} & \ldots & P_{1m} \\ P_{21} & \rho_2 P_2 & P_{23} & \ldots & P_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{m1} & P_{m2} & \ldots & P_{m(m-1)} & \rho_m P_m \end{bmatrix} \geq 0. \quad (21)$$

Lemma 7: Given m symmetric and positive-definite covariance matrices $P_1, P_2, \ldots, P_m \in \mathbb{R}^{n\times n}$ and a correlation coefficient bound $0 \leq \rho \leq 1$, consider the set of zero-mean RVs

$$\Omega = \{(a_1, a_2, \ldots, a_m) : \mathbf{E}\{a_i a_i^T\} = P_i, 1 \leq i \leq m; \mathbf{P}_m^{(\rho)} \geq 0\}.$$

Let $K_1, K_2, \ldots, K_m \in \mathbb{R}^{n\times n}$ be given nonsingular matrices and consider the RV

$$c = K_1 a_1 + K_2 a_2 + \cdots + K_m a_m \quad (22)$$

for any $(a_1, a_2, \ldots, a_m) \in \Omega$ and denote $P_c = \mathbf{E}\{cc^T\}$. Define

$$P(\delta_1, \delta_2, \ldots, \delta_m) = \sum_{i=1}^{m} w_i^{-1} K_i P_i K_i^T \quad (23)$$

with

$$w_i^{-1} = \delta_i^{-1}\rho_i + 1 - \rho_i \quad (24)$$

$\delta_i \in (0,\ 1),\ i = 1, 2, \ldots, m,$ subject to $\delta_1 + \delta_2 + \cdots + \delta_m = 1$. Then, for any $\delta_i, i = 1, 2, \ldots, m$, as defined above

$$P_c \leq P(\delta_1, \delta_2, \ldots, \delta_m),\ \ \forall (a_1, a_2, \ldots, a_m) \in \Omega.$$

Moreover, $\mathrm{Tr}(P(\delta_1,\delta_2,\ldots,\delta_m))$ is minimized by taking

$$\delta_i = (\rho_i \mathrm{Tr}(K_i P_i K_i^T))^{1/2} \left(\sum_{j=1}^{m} (\rho_j \mathrm{Tr}(K_j P_j K_j^T))^{1/2} \right)^{-1} \tag{25}$$

which results in

$$\min \mathrm{Tr}(P(\delta_1,\delta_2,\ldots,\delta_m)) = \sum_{i=1}^{m} (1-\rho_i)\mathrm{Tr}(K_i P_i K_i^T) + \left(\sum_{i=1}^{m} \sqrt{\rho_i \mathrm{Tr}(K_i P_i K_i^T)} \right)^2. \tag{26}$$

Proof: Denoting

$$\mathbf{P}_{(m-1),1} = [P_{m1}\ P_{m2}\ \ldots\ P_{m(m-1)}];$$
$$\mathbf{K}_1 = [K_1\ K_2\ \ldots K_{m-1}]$$

and using (21), we can express P_c as

$$\begin{aligned} P_c = & \sum_{i=1}^{m} (1-\rho_i) K_i P_i K_i^T \\ & + [\mathbf{K}_1 \mathbf{P}_{m-1}^{(\rho)} \mathbf{K}_1^T + K_m(\rho_m P_m) K_m^T \\ & + \mathbf{K}_1 \mathbf{P}_{(m-1),1}^T K_m^T + K_m \mathbf{P}_{(m-1),1} \mathbf{K}_1^T] \end{aligned} \tag{27}$$

where $\mathbf{P}_{m-1}^{(\rho)}$ is obtained from $\mathbf{P}_{m-1}^{(\rho)}$ by removing the last row and last column. We focus on the bracketed term above (the term inside $[\cdot]$). From (21), we know that

$$\mathbf{P}_{(m-1),1}^T = (\mathbf{P}_{m-1}^{(\rho)})^{1/2} F (\rho_m P_m)^{1/2}$$

for some F with $F^T F \le I$. As in the proof of Lemma 5, by applying the S-procedure [38] to (27), we obtain that

$$\begin{aligned} P_c \le & \sum_{i=1}^{m} (1-\rho_i) K_i P_i K_i^T + (1+\gamma_m^{-1}) K_m (\rho_m P_m) K_m^T \\ & + (1+\gamma_m) \mathbf{K}_1 \mathbf{P}_{m-1}^{(\rho)} \mathbf{K}_1^T \end{aligned} \tag{28}$$

for any $\gamma_m > 0$. Note that the term $\mathbf{K}_1 \mathbf{P}_{m-1}^{(\rho)} \mathbf{K}_1^T$ is the same as the bracketed term in (27) except that the dimension is reduced. We apply the derivation above repeatedly to yield

$$\begin{aligned} P_c \le & \sum_{i=1}^{m} (1-\rho_i) K_i P_i K_i^T \\ & + (1+\gamma_m^{-1}) K_m (\rho_m P_m) K_m^T \\ & + (1+\gamma_m)(1+\gamma_{m-1}^{-1}) K_{m-1} (\rho_{m-1} P_{m-1}) K_{m-1}^T \\ & + \ldots\ldots \\ & + (1+\gamma_m)\ldots(1+\gamma_2)(1+\gamma_1^{-1}) K_2 (\rho_2 P_2) K_2^T \\ & + (1+\gamma_m)\ldots(1+\gamma_2)(1+\gamma_1) K_1 (\rho_1 P_1) K_1^T \end{aligned} \tag{29}$$

for any $\gamma_1 > 0, \gamma_2 > 0, \ldots, \gamma_m > 0$. Take

$$\begin{aligned} \delta_m^{-1} &= (1+\gamma_m^{-1}) \\ \delta_{m-1}^{-1} &= (1+\gamma_m)(1+\gamma_{m-1}^{-1}) \\ &\ldots\ldots \\ \delta_2^{-1} &= (1+\gamma_m)\ldots(1+\gamma_2)(1+\gamma_1^{-1}) \\ \delta_1^{-1} &= (1+\gamma_m)\ldots(1+\gamma_2)(1+\gamma_1). \end{aligned}$$

It is clear that as all γ_i range from 0 to ∞, every δ_i ranges from 1 to 0. Also note that

$$(1+\gamma^{-1})^{-1} + (1+\gamma)^{-1} = 1$$

for any $\gamma > 0$, we obtain

$$(\delta_1+\delta_2)^{-1} = (1+\gamma_m)\ldots(1+\gamma_2).$$

Repeating the above, we obtain

$$(\delta_1+\delta_2+\delta_3)^{-1} = (1+\gamma_m)\ldots(1+\gamma_3)$$

etc. Keep repeating the above yields $\delta_1+\delta_2+\cdots+\delta_m = 1$. Therefore, the right-hand side of (29) equals $P(\delta_1,\delta_2,\ldots,\delta_m)$ in (23) if we take $w_i^{-1} = \delta_i^{-1}\rho_i + 1 - \rho_i$ for $i = 1,2,\ldots,m$.

It remains to show (25). We get from (23) that

$$\mathrm{Tr}(P(\delta_1,\delta_2,\ldots,\delta_m)) = \sum_{i=1}^{m} (\delta_i^{-1}\rho_i + (1-\rho_i)) \mathrm{Tr}(K_i P_i K_i^T).$$

Therefore, minimizing the above is the same as minimizing $\sum_{i=1}^{m} \delta_i^{-1} t_i$ subject to the given constraint of $\delta_1,\delta_2,\ldots,\delta_m$, where $t_i = \rho_i \mathrm{Tr}(K_i P_i K_i^T)$. This is easily given by the solution

$$\delta_i = \sqrt{t_i} \left(\sum_{j=1}^{m} \sqrt{t_j} \right)^{-1}$$

which is just (25).The expression (26) follows easily. ■

Remark 5: Although Lemma 7 generalizes Lemma 5, there is a main difference. First, Lemma 5 offers the optimal upper bound $\hat{P}$ of P_c with minimum trace. This turns out to be not possible for the general case of $m \ge 3$. This is due to the technical difficulty caused by the repeated use of S-procedure (see the proof of Lemma 7 for details). It is a well-known fact that repeated use of S-procedure cannot guarantee a tight upper bound in general, see, e.g., [40]. Hence, $P(\delta_1,\delta_2,\ldots,\delta_m)$ in (23) is only an upper bound of P_c in Lemma 7.

Next, we generalize Theorem 3 to multiple RVs.

Theorem 5: Given m symmetric and positive-definite covariances $P_1, P_2, \ldots, P_m \in \mathbb{R}^{n\times n}$, a mean value $\mu \in \mathbb{R}^n$ and a correlation coefficient bound (vector) $0 < \rho < 1$, consider the set of RVs

$$\Omega = \{(a_1,a_2,\ldots,a_m) :\ \mathbf{E}\{a_i\} = \mu, \mathrm{cov}(a_i) = P_i, \quad i = 1,2,\ldots,m; \mathbf{P}_m^{(\rho)} \ge 0\}. \tag{30}$$

For any nonsingular matrices $K_1, K_2, \ldots, K_m \in \mathbb{R}^{n\times n}$ subject to

$$K_1 + K_2 + \ldots + K_m = I \tag{31}$$

consider the unbiased linear combination

$$c = K_1 a_1 + K_2 a_2 + \cdots + K_m a_m \tag{32}$$

and its covariance upper bound $P(\delta_1, \delta_2, \ldots, \delta_m)$ for any $\delta_1, \delta_2, \ldots, \delta_m$ as defined in Lemma 7. Then, the minimum of $P(\delta_1, \delta_2, \ldots, \delta_m)$ is given by

$$\hat{P}(\delta_1, \delta_2, \ldots, \delta_m) = (w_1 P_1^{-1} + w_2 P_2^{-1} + \cdots + w_m P_m^{-1})^{-1} \tag{33}$$

with w_i defined in (24), i.e.

$$\hat{P}(\delta_1, \delta_2, \ldots, \delta_m) \leq P(\delta_1, \delta_2, \ldots, \delta_m)$$

for any nonsingular matrices $K_1, K_2, \ldots, K_m$ subject to (31). The minimum above is achieved by taking

$$K_i = w_i \hat{P}(\delta_1, \delta_2, \ldots, \delta_m) P_i^{-1}, \;\; i = 1, 2, \ldots, m. \tag{34}$$

Proof: The proof is extended from that of Theorem 3. Replace K_m with $I - K_1 - K_2 - \cdots - K_{m-1}$. Note that $P(\delta_1, \delta_2, \ldots, \delta_m)$ is convex in K_i, the minimum is obtained by differentiating $P(\delta_1, \delta_2, \ldots, \delta_m)$ with respect to K_i for $i = 1, 2, \ldots, m-1$ and setting them to zero. We obtain

$$w_i^{-1} K_i P_i - w_m^{-1} (I - K_1 - K_2 - \ldots - K_{m-1}) P_m = 0.$$

Defining

$$U = w_m^{-1} K_m P_m = w_m^{-1} (I - K_1 - K_2 - \cdots - K_{m-1}) P_m$$

we have $K_i = w_i U P_i^{-1}$ for $i = 1, 2, \ldots, m-1$. Also, we have $K_m = w_m U P_m^{-1}$ too. Substituting these solutions of K_i into $P(\delta_1, \delta_2, \ldots, \delta_m)$ gives

$$P(\delta_1, \delta_2, \ldots, \delta_m) = U \sum_{i=1}^{m} w_i P_i^{-1} U^T = U \hat{P}^{-1} U^T$$

with $\hat{P}$ in (33). Also, substituting these solutions of K_i into U gives

$$\begin{aligned} U &= w_m^{-1} (I - K_1 - K_2 - \cdots - K_{m-1}) P_m \\ &= w_m^{-1} (I - w_1 U P_1^{-1} - w_2 U P_2^{-1} - \cdots - w_{m-1} U P_{m-1}^{-1}) P_m \end{aligned}$$

which gives $U = \hat{P}(\delta_1, \delta_2, \ldots, \delta_m)$. Hence, the minimum of $P(\delta_1, \delta_2, \ldots, \delta_m)$ equals $\hat{P}(\delta_1, \delta_2, \ldots, \delta_m)$. Also, (34) is verified too. ■

Our next task is to optimize the parameters $\delta_1, \delta_2, \ldots, \delta_m$. That is, we need to solve the problem of

$$\begin{aligned} &\min \; \mathrm{Tr}(\hat{P}(\delta_1, \delta_2, \ldots, \delta_m)) \text{ subject to} \\ &\qquad (\delta_1, \delta_2, \ldots, \delta_m) \in D, \end{aligned} \tag{35}$$

where the constraint set

$$\begin{aligned} D = \{(\delta_1, \delta_2, \ldots, \delta_m) :\; &\delta_i \in (0,\; 1), i = 1, 2, \ldots, m; \\ &\delta_1 + \delta_2 + \cdots + \delta_m = 1\}. \end{aligned} \tag{36}$$

In other words, we need to generalize Lemma 6, Algorithm 1, and Theorem 4 to multiple RVs.

Lemma 8: Given symmetric and positive-definite matrices $P_1, P_2, \ldots, P_m$ and correlation coefficient bound (vector) $0 < \rho < 1$, consider the matrix function $\hat{P}(\delta_1, \delta_2, \ldots, \delta_m))$ as defined in (33). Then, $\mathrm{Tr}(\hat{P}(\delta_1, \delta_2, \ldots, \delta_m))$ is strictly convex over D in (36).

Proof: Due to the constraint that

$$\delta_m = 1 - \delta_1 - \delta_2 - \ldots - \delta_{m-1}$$

there are only $m-1$ free variables. Denoting $\delta = (\delta_1, \delta_2, \ldots, \delta_{m-1})$ and $\Delta = \hat{P}(\delta_1, \delta_2, \ldots, \delta_m)^{-1}$, we have (see proof of Lemma 6)

$$\begin{aligned} \frac{\partial^2}{\partial \delta^2} \mathrm{Tr}(\Delta^{-1}) = 2\mathrm{Tr} &\left(\Delta^{-1} \frac{\partial \Delta}{\partial \delta} \Delta^{-1} \frac{\partial \Delta}{\partial \delta} \Delta^{-1} \right) \\ &- \mathrm{Tr} \left(\Delta^{-1} \frac{\partial^2 \Delta}{\partial \delta^2} \Delta^{-1} \right). \end{aligned}$$

The first term is above nonnegative definite, so it suffices to show that the second term is positive definite over D. Indeed, for $i = 1, 2, \ldots, m-1$, it is easy to compute, using (24), that

$$\begin{aligned} \frac{\partial \Delta}{\partial \delta_i} &= \frac{1}{(\rho_i + (1-\rho_i)\delta_i)^2} P_i^{-1} \\ &\quad - \frac{1}{(\rho_m + (1-\rho_m)\delta_m)^2} P_m^{-1} \\ \frac{\partial^2 \Delta}{\partial \delta_i^2} &= -\frac{2(1-\rho_i)}{(\rho_i + (1-\rho_i)\delta_i)^3} P_i^{-1} \\ &\quad - \frac{2(1-\rho_m)}{(\rho_m + (1-\rho_m)\delta_m)^3} P_m^{-1} \\ \frac{\partial^2 \Delta}{\partial \delta_j \partial \delta_i} &= -\frac{2(1-\rho_m)}{(\rho_m + (1-\rho_m)\delta_m)^3} P_m^{-1}, \;\; j \neq i. \end{aligned}$$

This yields

$$\begin{aligned} &- \mathrm{Tr} \left(\Delta^{-1} \frac{\partial^2 \Delta}{\partial \delta^2} \Delta^{-1} \right) \\ &= \mathrm{diag} \left\{ \frac{2(1-\rho_i)}{(\rho_i + (1-\rho_i)\delta_i)^3} \mathrm{Tr}(\Delta^{-1} P_i^{-1} \Delta^{-1}) \right\} \\ &\quad + \left(\frac{2(1-\rho_m)}{(\rho_m + (1-\rho_m)\delta_m)^3} \mathrm{Tr}(\Delta^{-1} P_m^{-1} \Delta^{-1}) \right) ee^T \end{aligned}$$

where $e = [1 \; 1 \; \ldots \; 1]^T$. It is clear that the second term above is nonnegative definite and the first term is positive definite for all $0 < \rho < 1$. Hence, $\mathrm{Tr}(\hat{P}(\delta_1, \delta_2, \ldots, \delta_m))$ is strictly convex over D. ■

Following the spirit of Algorithm 1, we apply Lemma 8 to give an alternating projection algorithm for optimizing δ.

Algorithm 2: (Alternating Projection for Optimizing δ).

Step 1: Initialize $\delta_i = 1/m$ for $i = 1, 2, \ldots, m$;
Step 2: Compute w_i, $\hat{P}(\delta_1, \delta_2, \ldots, \delta_m)$ and K_i for $i = 1, 2, \ldots, m$ using (24), (33), and (34);
Step 3: Compute δ as in (25) and return to Step 2 until convergence.

The convergence property of the algorithm above is stated below. The proof is omitted as it is similar to that of Theorem 4.

Theorem 6: Given symmetric and positive-definite matrices $P_1, P_2, \ldots, P_m$ and correlation coefficient bound (vector) $0 < \rho < 1$, Algorithm 2 above convergences to the optimal solution for minimizing the trace of $\hat{P}(\delta_1, \delta_2, \ldots, \delta_m)$ in (33).

Remark 6: The alternating projection algorithm above is extremely efficient, as it does not involve any search (such as gradient search) in each iteration. Simulations show that only a few iterations (around 10) are sufficient. This observation of efficiency is in line with the common observations of alternating projection algorithms for semidefinite programming [41]. Note that our algorithm is an instance of such algorithms. In terms of the calculations required for each iteration, we note that the inverse matrices P_i^{-1} can be computed prior to the execution of the algorithm and they need to be done only once, and that calculations in (24), (25), and (34) are easy to do. The only more involved calculation is the matrix inversion $\hat{P}$ in (33) (the variables δ_i are suppressed for simplicity). When the matrix dimension n is large, efficient inversion algorithms for symmetric matrices [42] can be used.

Example 3: To demonstrate the obtained improvement by taking into account of the correlation coefficient bound, we consider the example with $m = n$ (recalling that n is the dimension of each RV) and

$$P_i = \text{diag}\{q, \ldots, q, 1, q, q, \ldots, q\}, i = 1, 2, \ldots, m$$

with $q > 1$, where the element 1 occurs at the ith diagonal location. For simplicity, we take ρ_i to be identical, i.e., $\rho_i = \rho_0$ for some constant $\rho_0, i = 1, 2, \ldots, m$.

Apply Theorem 5. Due to the symmetric structure of P_i, it is easy to see that the optimal $\delta_i = 1/m$, which in turn means that the optimal $w_i = w_0 = (m\rho_0 + 1 - \rho_0)^{-1}$. Subsequently, the optimal covariance bound (33) can be computed to be

$$\hat{P}(\delta_1, \delta_2, \ldots, \delta_m) = \text{diag}\left\{\frac{1 + (m-1)\rho_0}{1 + (m-1)q^{-1}}\right\} \tag{37}$$

and the corresponding

$$\text{Tr}(\hat{P}(\delta_1, \delta_2, \ldots, \delta_m)) = \frac{m + (m-1)m\rho_0}{1 + (m-1)q^{-1}}. \tag{38}$$

The resulting optimal K_i in (34) are given by

$$\begin{aligned} K_i &= w_i \hat{P}(\delta_1, \delta_2, \ldots, \delta_m) P_i^{-1} \\ &= \frac{1}{1 + (m-1)q^{-1}} \text{diag}\{q^{-1}, \ldots, q^{-1}, 1, q^{-1}, \ldots, q^{-1}\}. \end{aligned}$$

To see the advantage of considering partial correlations, we compare the bound in (37) with the relaxed Chebyshev centre CI (RCC-CI) algorithm and the information-theoretic fast CI (IT-FCI) algorithm in [26]. According to [26], both RCC-CI and IT-FCI are instances of the generalized CI, which has the following bound:

$$\hat{P}_{\text{wang}} = \left(\sum_{i=1}^{m} \delta_i P_i^{-1}\right)^{-1} \tag{39}$$

but each algorithm uses a different criterion to choose δ_i. Comparing with (33), it is clear that the above is the same as (33) with $\rho_i = \rho_0 = 1$. Since (38) uses the optimized δ_i (which is also valid for $\rho_0 = 1$), it follows that

$$\text{Tr}(\hat{P}_{\text{wang}}) \geq \frac{m + (m-1)m}{1 + (m-1)q^{-1}}$$

regardless how δ_i are chosen. In comparison with (38), we obtain

$$\frac{\text{Tr}(\hat{P}(\delta_1, \delta_2, \ldots, \delta_m))}{\text{Tr}(\hat{P}_{\text{wang}})} \leq \frac{m + (m-1)q_0}{2m - 1}.$$

The above clearly demonstrates the improvement by considering partial correlations.

Another interesting observation about this example is that, as $q \to \infty$, the optimal K_i becomes

$$K_i \to \text{diag}\{0, \ldots, 0, 1, 0, 0, \ldots, 0\}$$

which is independent of ρ_i. This is because in this case, the ith estimate (a_i) contains information in the ith component only, which implies that the optimal mixing gain matrices in (32) must be such that K_i "extract's" the ith component of a_i, regardless of the correlation coefficient bounds. However, we see from (38) that even in this case, the trace bound of $\hat{P}(\delta_1, \delta_2, \ldots, \delta_m)$ is significantly reduced by considering ρ_i.

V. Conclusion

We have generalized the standard covariance intersection algorithm to multiple RVs with any given correlation coefficient bound. For the case with two RVs, the resulting covariance is shown to be tight in the sense that it is the upper bound of the covariance, with the minimum trace, for the mixed RV. For the general case with more than two RVs, this optimality property is technically difficult to achieve and the resulting covariance is in general an upper bound of the covariance for the mixed RV. We have also provided an alternating projection algorithm to minimize the trace of the covariance by optimizing the associated scaling parameters. We expect that our results will make the already popular covariance intersection more applicable and more accurate for distributed estimation and information fusion problems.

References

[1] M. Reinhardt, B. Noack, and U. D. Hanebeck, "Closed-form optimization of covariance intersection for low-dimensional matrices," in *Proc. 15th Int. Conf. Inf. Fusion*, 2012, pp. 1891–1896.

[2] J. K. Uhlmann, "General data fusion for estimates with unknown cross covariances," *Proc. SPIE*, vol. 2755, 1996, Art. no. 536.

[3] S. J. Julier and J. K. Uhlmann, "Non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. 16th Amer. Control Conf.*, 1997, pp. 2369–2373.

[4] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[5] L. Chen, P. O. Arambel, and R. K. Mehra, "Estimation under unknown correlation: Covariance intersection revisited," *IEEE Trans. Automat. Control*, vol. 47, no. 11, pp. 1879–1882, Nov. 2002.

[6] P. Arambel, C. Rago, and R. Mehra, "Covariance intersection algorithm for distributed spacecraft state estimation," in *Proc. Amer. Control Conf.*, 2002, pp. 4398–4403.

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

[8] S. J. Julier and J. K. Uhlmann, "Using covariance intersection for SLAM," *Robot. Auton. Syst.*, vol. 55, no. 1, pp. 3–20, 2007.

[9] Y. Wang and X. Li, "A fast and fault-tolerant convex combination fusion algorithm under unknown cross-correlation," in *Proc.12th Int. Conf. Inf. Fusion*, 2009, pp. 571–578.

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

[11] Q. Guo, S. Chen, H. Leung, and S. Liu, "Covariance intersection based image fusion technique with application to pan sharpening in remote sensing," *Inf. Sci.*, vol. 180, no. 18, pp. 3434–3443, 2010.
[12] J. Sequeira, A. Tsourdos, and S. B. Lazarus, "Robust covariance estimation for data fusion from multiple sensors," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 12, pp. 3833–3844, Dec. 2011.
[13] Z. Deng, P. Zhang, W. Qi, J. Liu, and Y. Gao, "Sequential covariance intersection fusion Kalman filter," *Inf. Sci.*, vol. 189, pp. 293–309, 2012.
[14] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 891–902, Feb. 2012.
[15] H. Li, F. Nashashibi, and M. Yang, "Split covariance intersection filter: Theory and its application to vehicle localization," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1860–1871, Dec. 2013.
[16] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, 2013.
[17] J. K. Uhlmann, S. J. Julier, B. Kamgar-Parsi , M. O. Lanzagorta, and H.-J. S. Shyu, "NASA Mars rover: A testbed for evaluating applications of covariance intersection," *Proc. SPIE*, vol. 3693, pp. 140–149, Apr. 1999.
[18] W. Li, Z. Wang, G. Wei, L. Ma, J. Hu, and D. Ding, "A survey on multisensor fusion and consensus filtering for sensor networks," *Discrete Dyn. Nature Soc.*, vol. 2015, 2015, Art. no. 683701.
[19] K. C. Chang, C.-Y. Chong, and S. Mori, "Analytical and computational evaluation of scalable distributed fusion algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 4, pp. 2022–2034, Oct. 2010.
[20] S. J. Julier and J. K. Uhlmann, "General decentralized data fusion with covariance intersection (CI)," in *Handbook of Data Fusion*. Boca Raton, FL, USA: CRC Press, 2001.
[21] S. J. Julier, "Estimating and exploiting the degree of independent information in distributed data fusion," in *Proc. 12th Int. Conf. Inf. Fusion*, Seattle, WA, USA, Jul. 2009, pp. 772–779.
[22] M. B. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Proc. 5th Int. Conf. Inf. Fusion*, 2002, pp. 505–511.
[23] C. Y. Chong and S. Mori, "Convex combination and covariance intersection algorithms in distributed fusion," in *Proc. 4th Int. Conf. Inf. Fusion*, 2001.
[24] W. J. Farrell III and C. Ganesh, "Generalized Chernoff fusion approximation for practical distributed data fusion," in *Proc. 12th Int. Conf. Inf. Fusion*, 2009, pp. 555–562.
[25] J. Sijs and M. Lazar, "State fusion with unknown correlation: Ellipsoidal intersection," *Automatica*, vol. 48, no. 8, pp. 1874–1878, 2012.
[26] Y. Wang and X. R. Li, "Distributed estimation fusion with unavailable cross-correlation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 1, pp. 256–278, Jan. 2012.
[27] A. Riberio, I. D. Schizas, S. I. Roumeliotis, and G. B. Glannkis, "Kalman filtering in wireless sensor networks," *IEEE Control Syst. Mag.*, vol. 30, no. 2, pp. 66–96, Apr. 2010.
[28] B. Chen, W. A. Zhang, and L. Yu, "Distributed fusion estimation with missing measurements, random transmission delays and packet dropouts," *IEEE Trans. Automat. Control*, vol. 59, no. 7, pp. 1961–1967, Jul. 2014.
[29] B. Chen, G. Hu, D. W. C. Ho, and L. Yu, "Distributed covariance intersection fusion estimation for cyber-physical systems with communication constraints," *IEEE Trans. Automat. Control*, vol. 61, no. 12, pp. 4020–4026, Dec. 2016. doi: 10.1109/TAC.2016.2539221.
[30] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA, USA: Morgan Kaufman, 1988.
[31] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in gaussian graphical models of arbitrary topology," *Neural Comput.*, vol. 13, no. 10, pp. 2173–2200, 2001.
[32] D. E. Marelli and M. Fu, "Distributed weighted least-squares estimation with fast convergence for large-scale systems," *Automatica*, vol. 51, pp. 27–39, 2015.
[33] 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 *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, 2001, pp. 85–90.
[34] S. Reece and S. Roberts, "Robust, low-bandwidth, multi-vehicle mapping," in *Proc. 8th Int. Conf. Inf. Fusion*, 2005, pp. 1319–1326.
[35] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
[36] F. Bach, "Convex relaxations of structured matrix factorization," arXiv 1309.3117, pp. 1355–1362, 2013.
[37] T. Ando, "Concavity of certain maps on positive definite matrices and applications to Hadamard products," *Linear Algebr. Appl.*, vol. 26, pp. 203–241, 1979.
[38] V. A. Yakubovic, "S-procedure in nonlinear control theory," *Vestnik Leningrad Univ., Math.*, vol. 1, pp. 62–77, 1971.
[39] T. Iwasaki, G. Meinsma, and M. Fu, "Generalized S-procedure and finite freqency KYP lemma," *Math. Probl. Eng.*, vol. 6, pp. 305–320, 2000.
[40] G. Meinsma, T. Iwasaki, and M. Fu, "When is (D,G)-scaling both necessary and sufficient," *IEEE Trans. Automat. Control*, vol. 45, no. 9, pp. 1755–1759, Sep. 2000.
[41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[42] K. E. Atkinson, *An Introduction to Numerical Analysis*. Hoboken, NJ, USA: Wiley, 1989.

Zongze Wu received the B.S. degree in material forming and control, the M.S. degree in control science and engineering, and the Ph.D. degree in pattern reorganization and intelligence system, all from Xi'an Jiaotong University, Xi'an, China, in 1999, 2002, and 2005, respectively.

He is currently a Professor in the School of Automation, Guangdong University of Technology, Guangzhou, China. His research interests include automation control, signal processing, big data, and Internet of things. He has served as the Under-Secretary-General for Internet of Things and Information Technology Innovation Alliance in Guangdong, China.

Dr. Wu received the Microsoft Fellowship Award of the MSRA in 2003. He won the Technological Award first prize of Guangdong Province three times in 2008, 2013, and 2014, respectively. He got a second prize of Ministry of Education technological innovation twice, in 2012 and 2013.

Qianqian Cai received the B.S. degree in environmental engineering from the University of Shanghai for Science and Technology, Shanghai, China, in 2011, and the Ph.D. degree in engineering from the University of Newcastle, Callaghan, NSW, Australia, in 2016.

She is currently doing Postdoctoral research in the School of Automation, Guangdong University of Technology, Guangdong, China. Her research interests include water pollution control engineering, environmental monitoring, automation, and adaptive neural fuzzy inference systems.

Minyue Fu (F'04) received the Bachelor's degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 1982, and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, Madison, WI, USA, in 1983 and 1987, respectively.

From 1983 to 1987, he held a teaching assistantship and a research assistantship with the University of Wisconsin-Madison. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA. He joined the Department of Electrical and Computer Engineering, University of Newcastle, Callaghan, NSW, Australia, in 1989. He is currently a Chair Professor in Electrical Engineering and the Head of the School of Electrical Engineering and Computer Science. In addition, he was a Visiting Associate Professor with the University of Iowa in 1995–1996, and a Visiting Professor with Nanyang Technological University, Singapore, in 2002, a Changiang Professor with Shandong University in 2007–2010, a Qian-ren Scholar with Zhejiang University in 2010–2015, and a Bai-ren Scholar with the Guangdong University of Technology, China. His main research interests include control systems, and signal processing and communications.

Dr. Fu has been an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON SIGNAL PROCESSING, *Automatica*, and *Journal of Optimization and Engineering*.