Optimal Linear Fusion of Dimension-Reduced Estimates Using Eigenvalue Optimization

Robin Forsling, Zoran Sjanic, Fredrik Gustafsson and Gustaf Hendeby

Optimal Linear Fusion of Dimension-Reduced Estimates Using Eigenvalue Optimization

Robin Forsling*†, Zoran Sjanic*†, Fredrik Gustafsson*, and Gustaf Hendeby*
* Dept. of Electrical Engineering, Linköping University, Linköping, Sweden
e-mail: {firstname.lastname}@liu.se
† Saab AB, Linköping, Sweden
e-mail: {firstname.lastname}@saabgroup.com

***Abstract*—Data fusion in a communication constrained sensor network is considered. The problem is to reduce the dimensionality of the joint state estimate without significantly decreasing the estimation performance. A method based on scalar subspace projections is derived for this purpose. We consider the cases where the estimates to be fused are: (i) uncorrelated, and (ii) correlated. It is shown how the subspaces can be derived using eigenvalue optimization. In the uncorrelated case guarantees on mean square error optimality are provided. In the correlated case an iterative algorithm based on alternating minimization is proposed. The methods are analyzed using parametrized examples. A simulation evaluation shows that the proposed method performs well both for uncorrelated and correlated estimates.**

***Index Terms*—Communication constraints, dimension-reduced estimates, eigenvalue optimization, MSE optimal fusion, sensor networks, decentralized data fusion.**

I. Introduction

Estimation in a multi-sensor setting is a well-studied problem. A typical example is target tracking where multiple agents arranged as a sensor network estimate a common target. By utilizing multiple sensors the target estimate is improved in terms of accuracy [1]. The arrangement of sensors is often categorized after the flow and processing of information, *e.g.*, *centralized* and *decentralized* sensor networks (SN). In a centralized SN measurements are sent to a central agent for processing. In a decentralized SN measurements are preprocessed into estimates which are exchanged between agents. While a centralized arrangement offers the ability to compute *mean square error* (MSE) optimal estimates it suffers from high communication bandwidth requirements and high complexity. The decentralized arrangement on the other hand provides means for robust estimation and reduced communication burden, unfortunately at the expense of suboptimal estimates in the MSE sense [2]. It is worth noting that in certain cases also decentralized SN put high demands on the communication bandwidth [3].

An important aspect in SN is the cross-correlations between measurements and/or estimates. In a centralized SN cross-correlations are most often assumed zero. An MSE optimal estimator is in this case given by the *Kalman filter* (KF, [4]). This assumption does normally not hold in decentralized SN where nonzero cross-correlations are induced by preprocessing and communication of information [5]. If the cross-correlations are known the *Bar-Shalom-Campo* formula [6] defines an MSE optimal estimator. If tracking of all cross-correlations is not possible the *generalized information matrix filter* (GIMF, [7]) is an option. In the GIMF previously used information is explicitly subtracted in a decorrelation step. The *covariance intersection* (CI, [8]) algorithm is a popular conservative estimation method in cases when the cross-correlations are unknown since CI is robust to cross-correlations. In [9] a general approach based on robust optimization is proposed which is able to handle any sort of unknown or partially known cross-correlations in an efficient way.

This work has been supported by the Industry Excellence Center LINK-SIC funded by The Swedish Governmental Agency for Innovation Systems (VINNOVA) and Saab AB, and by the project Scalable Kalman filters funded by the Swedish Research Council (VR). G. Hendeby has received funding from the Center for Industrial Information Technology at Linköping University (CENIIT) grant no. 17.12.

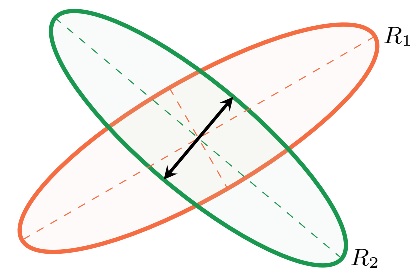

Fig. 1. Covariance ellipses of R_1 and R_2 corresponding to y_1 and y_2, respectively. An optimal scalar subspace to project y_2 onto is likely given by a vector approximately aligned with the minor axis of the R_2 ellipse, illustrated by the black arrow.

The problem of reducing the amount of exchanged data in a decentralized SN is studied in [10, 11]. This paper similarly deals with data fusion in an SN setting under communication constraints. The problem is to find an optimal way to reduce the estimate dimensionality. As an example, consider two estimates y_1 and y_2 of a common state x^0 in two dimensions, with covariances given by R_1 and R_2, respectively. The ellipses of R_1 and R_2 are defined in Fig. 1. Assume that it is only allowed to exchange y_2 projected along one direction. Moreover, assume that an estimate is to be derived from y_1 and the projected y_2. Then the best direction to project y_2 onto is likely given by a vector approximately aligned with the minor axis of the R_2 ellipse since in this direction y_2 is accurate while y_1 is inaccurate. In [11], the dimensionality of y_2 reduced by restricting y_2 to be mapped onto a combination of eigenvectors of R_2. In this paper that concept is extended to a mapping in an arbitrary direction. We consider the cases where: (i) y_1 and y_2 are uncorrelated, and (ii) y_1 and y_2 are correlated to an unknown degree.

A closely related problem is studied in [12, 13] where the computation of an optimal subspace to project measurements onto is solved with gradient descent. In [14–18] such problems in centralized and distributed configurations are studied where the proposed methods involve eigenvalue optimization similar to the proposed solution of this paper. However, the methodology proposed here is derived in a different way which results in a generalized eigenvalue problem. Moreover, a significant part of this paper is the focus on decentralized data fusion and estimates correlated to an unknown degree, where a novel algorithm is proposed for the latter.

This paper is organized as follows: The problem is stated in Sec. II. The proposed method for reducing the estimate dimensionality is described in Sec. III. In Sec. IV the proposed method is analyzed and in Sec. V it is further evaluated using a simulation study. Concluding remarks are given in Sec. VI.

II. Optimal Data Fusion of Reduced Estimates

The problem to reduce the dimensionality of one estimate is formulated as deriving a scalar subspace projection that yields optimal estimation performance given the reduced dimensionality. The performance depends on the specific fusion method used and the two applied fusion method are defined in the end of this section. We start with problem preliminaries.

A. Preliminaries

Consider estimation of an unknown state $x^0 \in \mathbb{R}^n$. Available are y_1 and y_2 modeled as

$$y_1 = x^0 + v_1, \qquad R_1 = \operatorname{cov}(v_1), \tag{1a}$$

$$y_2 = x^0 + v_2, \qquad R_2 = \operatorname{cov}(v_2), \tag{1b}$$

where v_1 and v_2 are noise, and R_1 and R_2 are the covariances of v_1 and v_2, respectively. It is assumed $\mathsf{E}\, v_1 = 0$ and $\mathsf{E}\, v_2 = 0$, where E is expectation operator, and that R_1 and R_2 are *positive definite* (PD). The estimate to be computed is denoted by $\hat{x}$ where P is the associated covariance. In this context y_1 and y_2 are interpreted as estimates but could also be measurements. Since y_1 and y_2 are estimates the considered problem is a fusion problem.

An estimate $\hat{x}$ of x^0 with covariance P is *conservative* if

$$P \succeq \mathsf{E}[(\hat{x} - x^0)(\hat{x} - x^0)^\mathsf{T}], \tag{2}$$

where $A \succeq B$ means that the difference $A - B$ is *positive semidefinite*. If $\hat{x}$ minimizes $\operatorname{tr}(P)$, then $\hat{x}$ is MSE optimal.

An *eigenvalue* $\lambda(A)$ of a matrix A is given by [19]

$$Au = \lambda u, \tag{3}$$

where u is an *eigenvector* associated with λ. If A is an $n \times n$ PD matrix, then the *eigenvalue problem* (EVP) in (3) has n solutions λ, not necessarily unique. The ith eigenvalue and an associated eigenvector form the pair (λ_i, u_i).

A *generalized eigenvalue* $\lambda(A, B)$ of matrices A and B is given by [19]

$$Ax = \lambda Bx, \tag{4}$$

where x is a *generalized eigenvector* associated with λ. If A and B are PD, then the *generalized eigenvalue problem* (GEVP) in (4) has n solutions λ, not necessarily unique. The ith generalized eigenvalue and an associated generalized eigenvector form the pair (λ_i, x_i).

The ellipsoid $\mathcal{E}(A)$ of an $n \times n$ PD matrix A is given by the set of points [20]

$$\mathcal{E}(A) = \left\{ z \in \mathbb{R}^n \mid z^\mathsf{T} A^{-1} z \leq 1 \right\}. \tag{5}$$

B. Problem Statement

The restriction to reduce the communication cost is that (y_2, R_2) is not globally accessible. Instead we have

$$y_\Psi = \Psi y_2 = \Psi(x^0 + v_2), \tag{6a}$$

$$R_\Psi = \operatorname{cov}(\Psi v_2) = \Psi R_2 \Psi^\mathsf{T}, \tag{6b}$$

where $\Psi \in \mathbb{R}^{1\times n}$ is non-zero and R_Ψ is the covariance of Ψv_2. The model in (6) is a mapping Ψ of y_2 to a 1D subspace, or equivalently, a projection of y_2 onto the subspace defined by the span of Ψ. A discussion about the communication gain is provided in Sec. V-C.

The problem is to find a row vector Ψ such that $\hat{x}$ derived from y_1 and y_Ψ is MSE optimal. An optimal Ψ is denoted by $\Psi^\star$. The actual value of P depends on the certain chosen estimator. Since an additional goal is that $\hat{x}$ is conservative the choice of estimator depends on the correlations between y_1 and y_Ψ. The two estimators to be used are described next.

C. Uncorrelated Estimates

If y_1 and y_Ψ are uncorrelated, then the measurement update of a KF can be applied to derive an MSE optimal estimate according to [21]

$$\hat{x} = P\left(R_1^{-1} y_1 + \Psi^\mathsf{T} R_\Psi^{-1} y_\Psi\right), \tag{7a}$$

$$P = \left(R_1^{-1} + \Psi^\mathsf{T} R_\Psi^{-1} \Psi\right)^{-1}. \tag{7b}$$

A more common version of P above is retrieved by applying the *matrix inversion lemma* [22] on (7b).

D. Correlated Estimates

If y_1 and y_Ψ are correlated, then CI [8] produces a conservative estimate irrespectively of the actual correlations between y_1 and y_Ψ. For the considered configuration CI is given by [23]

$$\hat{x} = P\left(\omega R_1^{-1} y_1 + (1-\omega)\Psi^\mathsf{T} R_\Psi^{-1} y_\Psi\right), \tag{8a}$$

$$P = \left(\omega R_1^{-1} + (1-\omega)\Psi^\mathsf{T} R_\Psi^{-1} \Psi\right)^{-1}, \tag{8b}$$

where in this case $\omega \in [0, 1]$ is found by minimizing $\operatorname{tr}(P)$. In [24] it is shown that CI is an MSE optimal conservative estimator for fusion of two estimates y_1 and y_2 of equal dimensions given that the correlations between y_1 and y_2 are completely arbitrary.

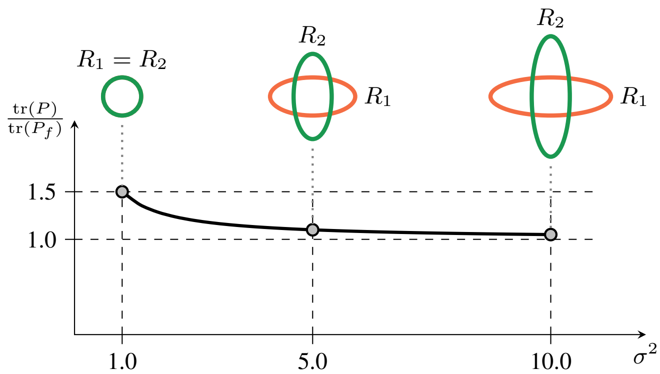

Fig. 2. The ratio $\frac{\mathrm{tr}(P)}{\mathrm{tr}(P_f)}$ indicates that using y_Ψ instead of y_2 does not significantly reduce the performance in many cases. The ellipses of R_1 and R_2 corresponding to $\sigma^2 = 1.0$, $\sigma^2 = 5.0$ and $\sigma^2 = 10.0$ are shown in the top.

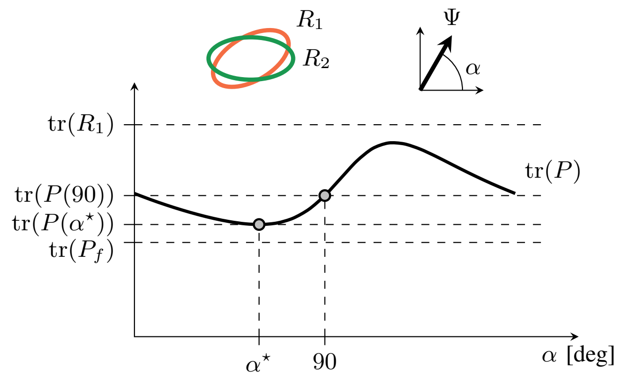

Fig. 3. The optimal Ψ is not trivially given by $u_{\min}^{\mathsf{T}} = \begin{bmatrix} 0 & 1 \end{bmatrix}$, where $u_{\min}$ is the eigenvector associated with $\lambda_{\min}(R_2)$. The parametrization of Ψ, and the R_1 and R_2 ellipses are shown in the top.

III. Selecting Ψ Using Eigenvalue Optimization

In this section it is shown how computing an optimal Ψ boils down to a GEVP. Without loss of generality Ψ is normalized. The case where y_1 and y_2 are uncorrelated is considered in Sec. III-B and in Sec III-C we move on to the correlated case. We start with motivating examples.

A. Motivation

First an example to see that it is not always a significant loss of performance to fuse y_1 and $y_\Psi = \Psi y_2$ instead of fusing y_1 and y_2. Assume y_1 and y_2 are uncorrelated. Let R_1, R_2 and Ψ be given by

$$R_1 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 1 & 0 \\ 0 & \sigma^2 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 1 & 0 \end{bmatrix},$$

where $\sigma^2 \geq 1$. Let P be given according to (7b) and

$$P_f = \left(R_1^{-1} + R_2^{-1}\right)^{-1},$$

which is (7b) with $\Psi = I$ and $R_\Psi = R_2$. We then have

$$\mathrm{tr}(P) = \frac{2\sigma^2 + 1}{\sigma^2 + 1}, \qquad \mathrm{tr}(P_f) = \frac{2\sigma^2}{\sigma^2 + 1}. \tag{9}$$

The ratio $\frac{\mathrm{tr}(P)}{\mathrm{tr}(P_f)} = 1 + \frac{1}{2\sigma^2}$ is plotted against σ^2 in Fig. 2, and as seen it approaches 1.0 relatively quickly. Hence, using y_Ψ instead of y_2 does not imply any significant degradation of performance as σ^2 grows. This motivates using y_Ψ and not y_2 in, *e.g.*, situations where the communication bandwidth is limited.

In the previous example Ψ was chosen to be an eigenvector associated with $\lambda_{\min}(R_2)$. Next we illustrate that Ψ cannot be derived in this way in general and still be expected to be optimal. Consider

$$R_1 = \begin{bmatrix} 3.2 & 1.2 \\ 1.2 & 1.8 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Psi = \begin{bmatrix} \cos\alpha & \sin\alpha \end{bmatrix}. \tag{10}$$

We have computed $\mathrm{tr}(P)$, with $P = P(\alpha)$ according to (7b) and R_1 and $R_\Psi = \Psi R_2 \Psi^{\mathsf{T}}$ according to above, for different values of α. The results are shown in Fig. 3. It is seen that $\Psi^\star$ is not trivially given by $u_{\min}^{\mathsf{T}} = \begin{bmatrix} 0 & 1 \end{bmatrix}$. Instead $\Psi^\star = \begin{bmatrix} \cos\alpha^\star & \sin\alpha^\star \end{bmatrix}$, where $\alpha^\star \approx 59°$. The ratio between the maximum and minimum value of $\mathrm{tr}(P)$ is $\frac{\max \mathrm{tr}(P)}{\min \mathrm{tr}(P)} \approx 1.73$.

B. Uncorrelated Estimates

Consider $(\hat{x}, P)$ according to (7). As the goal is to derive Ψ such that $\hat{x}$ is MSE optimal we want to minimize

$$\begin{aligned} \mathrm{tr}(P) &= \mathrm{tr}\left(\left(R_1^{-1} + \Psi^{\mathsf{T}} R_\Psi^{-1} \Psi\right)^{-1}\right) \\ &= \mathrm{tr}\left(R_1 - R_1 \Psi^{\mathsf{T}} \left(R_\Psi + \Psi R_1 \Psi^{\mathsf{T}}\right)^{-1} \Psi R_1\right) \\ &= \mathrm{tr}(R_1) - \mathrm{tr}\left(R_1 \Psi^{\mathsf{T}} \left(\Psi R_1 \Psi^{\mathsf{T}} + \Psi R_2 \Psi^{\mathsf{T}}\right)^{-1} \Psi R_1\right), \end{aligned}$$

where the matrix inversion lemma, the linear property of trace and $R_\Psi = \Psi R_2 \Psi^{\mathsf{T}}$ have been used. Since R_1 is constant and $\Psi \in \mathbb{R}^{1 \times n}$, minimization of $\mathrm{tr}(P)$ reduces to maximization of

$$\mathrm{tr}\left(\frac{R_1 \Psi^{\mathsf{T}} \Psi R_1}{\Psi(R_1 + R_2)\Psi^{\mathsf{T}}}\right).$$

Using the cyclic property of trace we have

$$\begin{aligned} \mathrm{tr}\left(\frac{R_1 \Psi^{\mathsf{T}} \Psi R_1}{\Psi(R_1 + R_2)\Psi^{\mathsf{T}}}\right) &= \mathrm{tr}\left(\frac{\Psi R_1^2 \Psi^{\mathsf{T}}}{\Psi(R_1 + R_2)\Psi^{\mathsf{T}}}\right) \\ &= \frac{\Psi R_1^2 \Psi^{\mathsf{T}}}{\Psi(R_1 + R_2)\Psi^{\mathsf{T}}}, \end{aligned}$$

where in the last step it has been used that $\mathrm{tr}(a) = a$ for a scalar a. We end up with

$$\underset{\Psi \neq 0}{\text{maximize}} \quad \frac{\Psi R_1^2 \Psi^{\mathsf{T}}}{\Psi(R_1 + R_2)\Psi^{\mathsf{T}}},$$

or equivalently

$$\underset{\Psi \neq 0}{\text{minimize}} \quad \frac{\Psi(R_1 + R_2)\Psi^{\mathsf{T}}}{\Psi R_1^2 \Psi^{\mathsf{T}}},$$

A more familiar notation is retrieved by letting $x = \Psi^{\mathsf{T}}$, $Q = R_1^2$ and $S = R_1 + R_2$, *i.e.*,

$$\underset{x \neq 0}{\text{minimize}} \quad \frac{x^{\mathsf{T}} S x}{x^{\mathsf{T}} Q x}, \tag{11}$$

where the function to be minimized is known as the (generalized) *Rayleigh quotient* [19]. Let $x = aw$ where $a \neq 0$ is a scalar and w is a vector which is parallel to x. We then have

$$\frac{x^\mathsf{T} S x}{x^\mathsf{T} Q x} = \frac{(aw)^\mathsf{T} S aw}{(aw)^\mathsf{T} Q aw} = \frac{a^2 (w^\mathsf{T} S w)}{a^2 (w^\mathsf{T} Q w)} = \frac{w^\mathsf{T} S w}{w^\mathsf{T} Q w},$$

which implies that if x solves (11) then so does $w = \frac{x}{a}$ for any $a \neq 0$. This means that only the direction of x is of interest. Since $Q = R_1^2$ where R_1 is invertible it is possible to define $x = R_1^{-1} z$, where $z \in \mathbb{R}^n$. Moreover, since we are only interested in the direction of x, or equivalently of z, we can without loss of generality impose the constraint $z^\mathsf{T} z = x^\mathsf{T} Q x = 1$. Instead of (11) we therefore consider the problem [25]

$$\begin{aligned} \underset{x \neq 0}{\text{minimize}} \quad & x^\mathsf{T} S x \\ \text{subject to} \quad & x^\mathsf{T} Q x = 1. \end{aligned} \tag{12}$$

The Lagrangian of (12) is

$$\mathcal{L} = x^\mathsf{T} S x - \lambda (x^\mathsf{T} Q x - 1), \tag{13}$$

where λ is a *Lagrange multiplier* associated with the constraint [20]. To solve (12) we differentiate $\mathcal{L}$ w.r.t. x and set the result equal to zero. This yields

$$\frac{\partial \mathcal{L}}{\partial x} = 2Sx - 2\lambda Q x = 0 \implies Sx = \lambda Q x, \tag{14}$$

which is a GEVP, *cf.* (4). The solution to (11) is derived by solving $Sx = \lambda Q x$ and choosing $(\lambda_{\min}, x_{\min})$. The sought mapping is given by $\Psi^\star = x_{\min}^\mathsf{T} / \|x_{\min}\|$, where normalization is not necessary but is done for convenience. This method of selecting Ψ is denoted the *generalized eigenvalue optimization* (GEVO) method.

1) Example of the Generalized Eigenvalue Problem: To illustrate the considered GEVP, assume R_1 and R_2 are given by (10). Let $Q = R_1^2$ and $S = R_1 + R_2$. The equation

$$Sx = \lambda Q x, \tag{15}$$

now has two solutions $(\lambda_{\min}, x_{\min})$ and $(\lambda_{\max}, x_{\max})$. The results are illustrated in Fig. 4.

2) Recovering an Eigenvalue Problem: We will now see that it is also possible to formulate the problem as an EVP. Recall that $Q = R_1^2$ and $S = R_1 + R_2$. A variable substitution $z = R_1 x$ transforms the problem in (11) into

$$\underset{z \neq 0}{\text{minimize}} \quad \frac{z^\mathsf{T} (R_1^{-1} + R_1^{-1} R_2 R_1^{-1}) z}{z^\mathsf{T} z}. \tag{16}$$

Since the ratio in (16) is invariant under scaling of z, an equivalent problem is

$$\begin{aligned} \underset{z \neq 0}{\text{minimize}} \quad & z^\mathsf{T} (R_1^{-1} + R_1^{-1} R_2 R_1^{-1}) z \\ & z^\mathsf{T} z = 1. \end{aligned} \tag{17}$$

The Lagrangian is in this case given by

$$\mathcal{L} = z^\mathsf{T} (R_1^{-1} + R_1^{-1} R_2 R_1^{-1}) z - \lambda (z^\mathsf{T} z - 1). \tag{18}$$

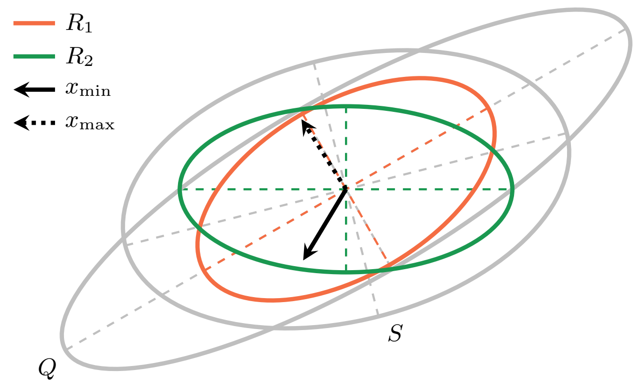

Fig. 4. Illustration of the generalized eigenvalue problem $Sx = \lambda Q x$ where $Q = R_1^2$ and $S = R_1 + R_2$. The ellipses of R_1, R_2, Q and S are provided along with their principle axes represented by dashed lines. The generalized eigenvectors $x_{\min}$ and $x_{\max}$ correspond to $\lambda_{\min}(S, Q)$ and $\lambda_{\max}(S, Q)$, respectively.

Differentiating $\mathcal{L}$ w.r.t. z and equating the result to zero yield

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial z} &= 2(R_1^{-1} + R_1^{-1} R_2 R_1^{-1}) z - 2\lambda z = 0 \\ &\implies (R_1^{-1} + R_1^{-1} R_2 R_1^{-1}) z = \lambda z, \end{aligned} \tag{19}$$

which is an EVP, *cf.* (3), for the matrix $R_1^{-1} + R_1^{-1} R_2 R_1^{-1}$. After finding the eigenvector $z_{\min}$ corresponding to $\lambda_{\min}(R_1^{-1} + R_1^{-1} R_2 R_1^{-1})$, the sought projection is given by $\Psi^\star = (R_1^{-1} z_{\min})^\mathsf{T} / \|R_1^{-1} z_{\min}\|$.

C. Correlated Estimates

To illustrate the behavior when y_1 and y_2 are correlated, and the usage of CI formulas in (8) instead of KF, consider R_1, R_2 and Ψ defined as in (10). Let P^{KF} and P^{CI} be given according to (7b) and (8b), respectively. Let

$$\begin{aligned} P_f^{\text{KF}} &= \left(R_1^{-1} + R_2^{-1}\right)^{-1}, \\ P_f^{\text{CI}} &= \left(\omega R_1^{-1} + (1-\omega) R_2^{-1}\right)^{-1}, \end{aligned}$$

where ω is found by minimizing $\operatorname{tr}(P_f^{\text{CI}})$. The results of varying $\alpha \in [0°, 180°]$ are illustrated in Fig. 5. The angle α that defines the optimal Ψ differs when using (8b) compared to (7b), and is given by $\alpha_{\text{CI}}^\star \approx 66°$ and $\alpha_{\text{KF}}^\star \approx 59°$ in case of CI and KF, respectively. By examining $\operatorname{tr}(P_f^{\text{CI}})$ and $\operatorname{tr}(P_f^{\text{KF}})$ we see that the gain from using CI is low. This is due to the fact that the geometry corresponding to R_1 and R_2 is not well-suited for this estimator.

Based on the observations above we conclude that if y_1 and y_Ψ are correlated and are fused using CI, then the optimal Ψ should be derived by other means than using the GEVO method for uncorrelated estimates. In Algorithm 1 an iterative approach based on *alternating minimization*[1] [26] and CI is provided.

[1]Also known as *block coordinate descent* [26].

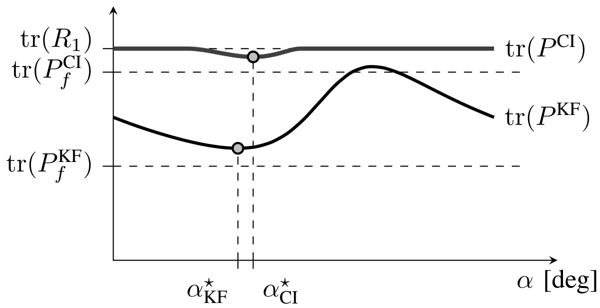

Fig. 5. Motivation for not using the GEVO method derived for uncorrelated y_1 and y_2 in case they actually are correlated. The optimal Ψ for CI deviates from the KF case.

Algorithm 1 The GEVO Method For Correlated Estimates

Input: R_1 and (y_2, R_2)

1: **Initialization**. Let $i = 0$. Compute ω_0 by solving

$$\underset{\omega}{\text{minimize}} \quad \text{tr}\left((\omega R_1^{-1} + (1-\omega)R_2^{-1})^{-1}\right).$$

2: Let $i \leftarrow i+1$. Compute $x_{\min}$ by solving

$$\begin{aligned}&\underset{\lambda,x}{\text{minimize}} && \lambda \\ &\text{subject to} && Sx = \lambda Qx,\end{aligned}$$

where $Q = R_1^2/\omega_{i-1}^2$ and $S = R_1/\omega_{i-1} + R_2/(1-\omega_{i-1})$.

3: Let $\Psi = x_{\min}^\mathsf{T}/\|x_{\min}\|$. Compute ω_i by solving

$$\underset{\omega}{\text{minimize}} \quad \text{tr}\left((\omega R_1^{-1} + (1-\omega)\Psi^\mathsf{T}(\Psi R_2 \Psi^\mathsf{T})^{-1}\Psi)^{-1}\right).$$

4: If $|\omega_i - \omega_{i-1}| < \epsilon$ or if the maximum number of iterations has been reached, then terminate with most recently computed Ψ and $y_\Psi = \Psi y_2, R_\Psi = \Psi R_2 \Psi^\mathsf{T}$. Otherwise go back to step 2.

Output: (Ψ, y_Ψ, R_Ψ)

IV. Problem Intuition And Method Analysis

In this section the GEVO method is analyzed. The goal is to give intuition for the behavior under different circumstances. The GEVO method for uncorrelated estimates are analyzed using a parametrized problem and compared to selecting Ψ based on computing $\lambda_{\min}(R_2)$. Finally, The GEVO method for correlated estimates is analyzed.

A. The GEVO Method For Uncorrelated Estimates

An analysis of the GEVO method used for deriving $\Psi^\star$ is now acquired by considering a parametrized but restricted example. Let R_1 and R_2 be parametrized as

$$R_1 = T(\phi)\begin{bmatrix}\sigma^2 & 0 \\ 0 & 1\end{bmatrix}T(\phi)^\mathsf{T}, \qquad R_2 = \begin{bmatrix}2\varsigma^2 & 0 \\ 0 & \frac{2}{\varsigma^2}\end{bmatrix}, \tag{20}$$

where the rotation matrix $T(\phi)$ is defined as

$$T(\phi) = \begin{bmatrix}\cos\phi & -\sin\phi \\ \sin\phi & \cos\phi\end{bmatrix}.$$

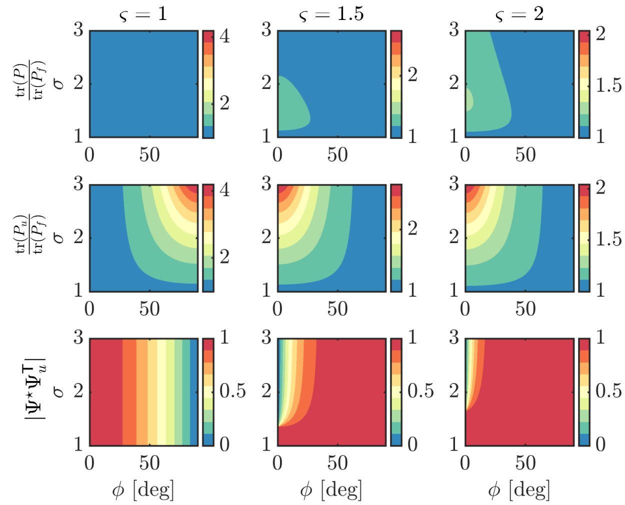

Fig. 6. Results of the GEVO method for the problem parametrized in $(\phi, \sigma, \varsigma)$. The top, middle and bottom row correspond to evaluation of $\frac{\text{tr}(P)}{\text{tr}(P_f)}$, $\frac{\text{tr}(P_u)}{\text{tr}(P_f)}$ and $|\Psi^\star\Psi_u^\mathsf{T}|$, respectively. Each column is for one value of ς.

In the next ϕ and σ are varied for each ς. For each configuration of $(\phi, \sigma, \varsigma)$ the following are computed

$$\begin{aligned}P_f &= \left(R_1^{-1} + R_2^{-1}\right)^{-1}, \\ P &= \left(R_1^{-1} + (\Psi^\star)^\mathsf{T}(\Psi^\star R_2 (\Psi^\star)^\mathsf{T})^{-1}\Psi^\star\right)^{-1}, \\ P_u &= \left(R_1^{-1} + \Psi_u^\mathsf{T}(\Psi_u R_2 \Psi_u^\mathsf{T})^{-1}\Psi_u\right)^{-1},\end{aligned}$$

with $\Psi^\star = x_{\min}^\mathsf{T}/\|x_{\min}\|$ and $\Psi_u = u_{\min}^\mathsf{T}/\|u_{\min}\|$, where $x_{\min}$ is a generalized eigenvector associated with $\lambda_{\min}(S, Q)$ and $u_{\min}$ is an eigenvector associated with $\lambda_{\min}(R_2)$.

The results are illustrated in Fig. 6, where the ratios $\frac{\text{tr}(P)}{\text{tr}(P_f)} \geq 1$ and $\frac{\text{tr}(P_u)}{\text{tr}(P_f)} \geq 1$ have been computed to illustrate the performance when using $\Psi^\star R_2 (\Psi^\star)^\mathsf{T}$ and $\Psi_u R_2 \Psi_u^\mathsf{T}$ instead of R_2. The absolute value of $\Psi^\star\Psi_u^\mathsf{T}$ is included to visualize the difference between the selected $\Psi^\star$ and Ψ_u.

For most of the configurations of $(\phi, \sigma, \varsigma)$, $\text{tr}(P)$ is close to $\text{tr}(P_f)$ which means that the performance of the GEVO method is close to the performance of optimal fusion of (y_1, R_1) and (y_2, R_2). It can also be concluded that while Ψ_u only deviate by a small amount from $\Psi^\star$, as indicated by $|\Psi^\star\Psi_u^\mathsf{T}|$ being close to 1, there can still be a significant degradation in performance when using Ψ_u instead of $\Psi^\star$.

B. The GEVO Method For Correlated Estimates

A similar analysis as above is done for the correlated case with P according to (8b). The same parametrization is used for R_1 and R_2 as in (20), but now $\sigma = 2.5$ is fixed and ϕ is varied over several values. The results are illustrated in Fig. 7, where each plot shows $\text{tr}(P)$ as a function of ω and α. Each row and column correspond to one value of ς and ϕ, respectively. The red cross corresponds to an optimal Ψ found by a grid search and the black plus sign corresponds to that Ψ computed by Algorithm 1. The dashed black line illustrates ω given by step 1 of Algorithm 1.

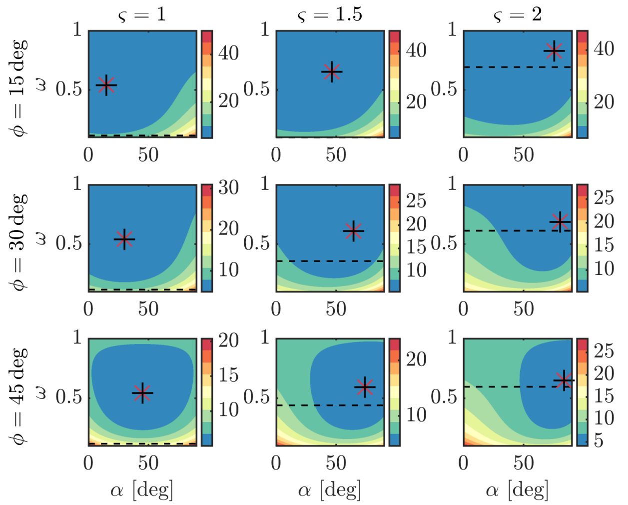

Fig. 7. Each row and column correspond to one value of ς and ϕ, respectively. The red cross and the black plus sign correspond to an optimal Ψ found by a grid search and as computed by Algorithm 1, respectively. The dashed black line resembles ω given by step 1 of Algorithm 1.

In each plot the solution computed by Algorithm 1 coincides with the solution of a grid search. The dashed line is always below ω associated with the optimal Ψ since $R_2^{-1} \succeq \Psi^\mathsf{T}(\Psi R_2 \Psi^\mathsf{T})^{-1}\Psi$ CI should favour R_1^{-1} more if R_2^{-1} is replaced by $\Psi^\mathsf{T}(\Psi R_2 \Psi^\mathsf{T})^{-1}\Psi$, *cf.* (8b). This also indicates that it is important to not just solve for $\Psi^\star$ by naively using a fixed ω, *e.g.*, ω as computed in step 1 of Algorithm 1.

V. Experimental Evaluation

To further analyze the performance of the GEVO method a *Monte Carlo* (MC) based simulation study is done. The purpose of this experimental evaluation is to investigate the performance decreases when using (y_Ψ, R_Ψ) instead of (y_2, R_2) to reduce the communication cost in a typical target tracking scenario.

A. Evaluation Metrics

Root mean square error (RMSE) of the position components is used to evaluate the performance. Since for conservative estimators the RMSE is generally smaller than the square root of the trace of the associated covariance P_{pos}, where P_{pos} denotes the covariance of the position components, the RMSE curves are accompanied by $\sqrt{\operatorname{tr}(P_{\text{pos}})}$.

To evaluate if the computed estimates are conservative the *average normalized estimation error squared* (ANEES, [27]) metric is used. Let M denote the number of MC runs. Let $\hat{x}^i_{k|k}$ be a local estimate of the state x^0_k at time k in the ith MC run. ANEES ε_k at time k is computed as

$$\varepsilon_k = \frac{1}{nM}\sum_{i=1}^{M}\varepsilon^i_k, \tag{21a}$$

$$\varepsilon^i_k = (\hat{x}^i_{k|k} - x^0_k)^\mathsf{T}(P^i_{k|k})^{-1}(\hat{x}^i_{k|k} - x^0_k), \tag{21b}$$

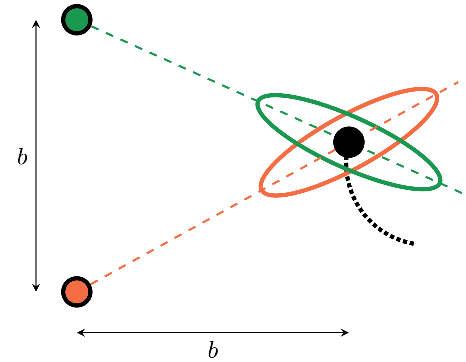

Fig. 8. Simulated scenario. The green and the orange circles are stationary agents that track a common target (black circle) by observing, filtering and exchanging estimates of the target. The ellipses centered at the target represents 95% confidence ellipses of the measurement covariance R^S_k in each agent. The target moves along the dotted curve.

where $P^i_{k|k}$ is the covariance computed for $\hat{x}^i_{k|k}$. If ANEES of an estimator is significantly higher than 1, then this estimator is too optimistic and not conservative. If ANEES of an estimator is significantly lower than 1, then this estimator is overly conservative. It is desirable to have an estimator with ANEES being within a confidence interval centered around approximately 1. If this is not possible an estimator should strive for $\varepsilon_k < 1$.

Let $z_k = h(x^0_k) + e_k$ be a given measurement model, where $h(x)$ is the measurement function which maps from state coordinates to measurement coordinates, and $R^\mathrm{S}_k = \operatorname{cov}(e_k)$ is the covariance of the measurement noise e_k. Since x^0_k is fixed throughout all MC runs the parametric *Cramér-Rao lower bound* (CRLB, [28]) P^0 can be computed recursively according to [29]

$$P^0_{k|k} = \left((P^0_{k|k-1})^{-1} + (H^0_k)^\mathsf{T}(R^\mathrm{S}_k)^{-1}H^0_k\right)^{-1}, \tag{22a}$$

$$P^0_{k+1|k} = F^0_k P^0_{k|k}(F^0_k)^\mathsf{T} + Q^0_k, \tag{22b}$$

where $H^0_k = \left.\frac{\partial h(x)}{\partial x}\right|_{x=x^0_k}$, F^0_k is an assumed process model for x^0_k and Q^0_k is the process noise covariance derived from the true dynamics of x^0_k.

B. Simulation Specifications

The considered scenario is defined in two spatial dimensions and is shown in Fig. 8. Two stationary agents illustrated by green and orange circles track a dynamic target which is illustrated by the black circle and that follows the dotted path. The sampling time is $T_s = 1$ s. At every k, a sensor in each agent generates a new measurement in polar coordinates of the target. The standard deviation of the radial noise is $\sigma_r \approx \frac{b}{7}$, where b is the baseline as defined in Fig. 8. The standard deviation of the azimuthal noise is $\sigma_\theta = 2^\circ$. In Fig. 8 we have also visualized 95% confidence ellipses of R^S_k transformed into state coordinates. The ratio between the uncertainty in the radial and azimuthal directions is $\frac{\sigma_r}{r\sigma_\theta} \approx 4$, where r is the approximate distance to the target.

The measurements are filtered using an *extended Kalman filter* (EKF, [30]) and a *constant acceleration model* (CAM, [31]) is used to capture the dynamics of x_k^0. Hence $n = 6$ and $x_k^0 \in \mathbb{R}^6$. The process noise covariance is tuned such that a *local Kalman filter* (LKF, see below) initially reaches ANEES of approximately 1. The local estimates are exchanged at a rate of $\frac{1}{2T_s}$ over a datalink for fusion at the other agent. Selection of the exchanged data is based on the GEVO method. Curves for the following methods and quantities are computed for evaluation:

CRLB The Cramér-Rao lower bound as given by the recursions above.

LKF A local filter that only uses local information and no communicated estimates.

GEVO-KF The GEVO method for uncorrelated estimates, see Sec. III-B.

GEVO-KF-P $\sqrt{\text{tr}(P_{\text{pos}})}$ for GEVO-KF.

GEVO-CI The GEVO method for correlated estimates, see Sec. III-C.

GEVO-CI-P $\sqrt{\text{tr}(P_{\text{pos}})}$ for GEVO-CI.

C. Communication Considerations

An $n \times n$ covariance matrix R_2 consists of $\frac{n(n+1)}{2}$ independent parameters. When exchanging a full estimate $y_2 \in \mathbb{R}^n$ and covariance R_2 we hence need to transmit $N_{\text{full}} = n + \frac{n(n+1)}{2} = \frac{n(n+3)}{2}$ parameters. A primary motivation for the dimension-reduced estimates is to decrease the communication cost. Since $\Psi \in \mathbb{R}^{1\times n}$ both $y_\Psi = \Psi y_2$ and $R_\Psi = \Psi R_2 \Psi^\mathsf{T}$ are scalars, but as an agent receiving (y_Ψ, R_Ψ) also requires knowledge about Ψ we have to encode Ψ in the transmitted message somehow. This is accomplished as follows.

Let $\Psi \in \mathbb{R}^{1\times n}$ be of unit length. The transmitting agent then sends $(y_\Psi, R_\Psi \Psi)$ where the product $R_\Psi \Psi$ encodes both R_Ψ and Ψ. Upon reception of $(y_\Psi, R_\Psi \Psi)$ the receiving agent extracts R_Ψ and Ψ using the fact that $R_\Psi > 0$ and $\|\Psi\| = 1$. When transmitting $(y_\Psi, R_\Psi \Psi)$ the number of transmitted parameters is reduced to $N_{\text{red}} = n + 1$. We hence have the ratio $\frac{N_{\text{red}}}{N_{\text{full}}} = \frac{2(n+1)}{n(n+3)}$ which for large n is approximately given by $\frac{2}{n}$. This ratio is directly related to the reduced communication cost and is strictly less than one for all $n > 1$. For example, assuming three spatial coordinates and a CAM such that $n = 9$ results in $\frac{N_{\text{red}}}{N_{\text{full}}} \approx 19\%$. Hence for $n = 9$, transmitting $(y_\Psi, R_\Psi \Psi)$ instead of (y_2, R_2) means a communication reduction of more than 80%. The reduced communication cost is more generally illustrated in Fig. 9.

In a real-world case, *e.g.*, a scenario similar to the simulated one, each agent only has access to local information. This means that when computing an optimal Ψ, R_1 is not accessible. In the simulations R_1 is however assumed globally known. To maintain uncorrelated estimates, as is necessary in the GEVO-KF case, previously used information is explicitly subtracted. This decorrelation procedure is done locally when selecting $(y_\Psi, R_\Psi \Psi)$ for communication to the other agent.

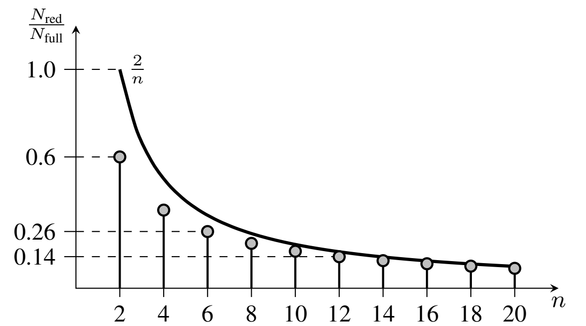

Fig. 9. Reduced communication cost illustrated by stem plots of $\frac{N_{\text{red}}}{N_{\text{full}}}$ as a function of n. The ratio $\frac{N_{\text{red}}}{N_{\text{full}}}$ asymptotically behaves as $\frac{2}{n}$.

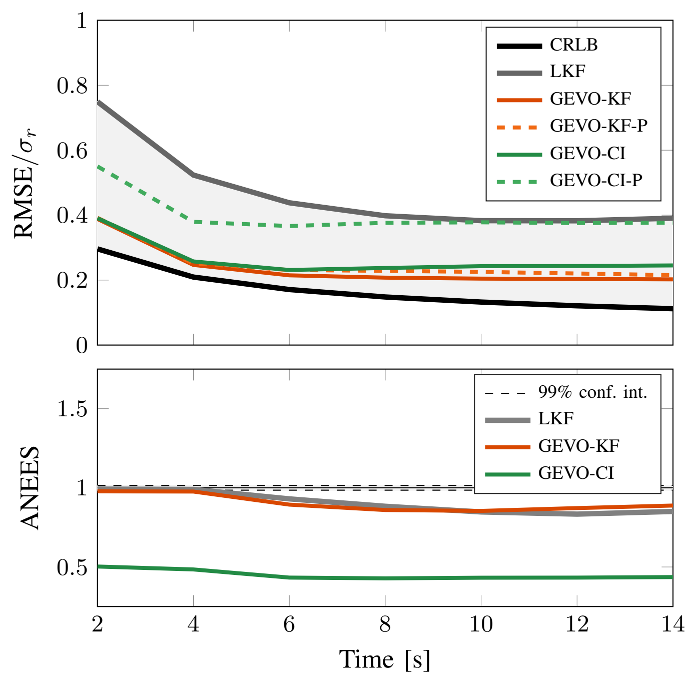

Fig. 10. *Top:* Position RMSE normalized with σ_r. *Bottom:* ANEES including a 99% confidence interval around 1.

D. Results

The estimation results of the upper agent (green) are shown in Fig. 10 for $M = 10\,000$ MC runs. The results of the other agent is similar. In the RMSE plot it is seen that the RMSE when using the GEVO method for selection of Ψ is relatively close to CRLB, despite that y_2 has been projected down to a scalar quantity $y_\Psi = \Psi y_2$. GEVO-KF performs slightly better than GEVO-CI w.r.t. RMSE. When it comes to $\sqrt{\text{tr}(P_{\text{pos}})}$ GEVO-KF is superior GEVO-CI, see the curves for GEVO-KF-P and GEVO-CI-P. This is because CI is overly conservative, as is also indicated by the ANEES plot. The ANEES plot also suggests that all methods are conservative. A 99% confidence interval for ANEES is also provided.

VI. Conclusions

We have considered the problem of finding optimal projections, defined by Ψ, that map a measurement vector or estimate onto a scalar subspace. The *generalized eigenvalue optimization* (GEVO) method has been proposed for computation of an optimal Ψ. In case of uncorrelated measurements and/or estimates the GEVO method guarantees mean square error optimal estimates. In case of correlated measurements and/or estimates the GEVO method and covariance intersection have been combined into an alternating minimization algorithm. A method analysis and an experimental evaluation have been done that show that the GEVO method and the alternating minimization algorithm perform well in various problems and geometries.

The main application that has been considered for the proposed methods in this paper is fusion of two estimates. More generally we have distinguished the proposed methods as relevant in the following applications:

(i) *Communication efficient data fusion*. To reduce the communication bandwidth, redundant parts of measurements and estimates can be removed before being exchanged.

(ii) *System design*. When designing, *e.g.*, a sensor network, optimal coordinates for sensor placement can be derived.

(iii) *Experimental design*. The GEVO method can be used to design experiments involving sensor placements. In this case y_1 might be considered to be a prior.

In this paper we have only considered the problem where Ψ is a row vector. Possible future work includes extending the GEVO method to allow for matrix valued Ψ. A convergence analysis of the alternating minimization algorithm is also highly relevant. The case when perfect knowledge about R_1 and R_2 is not available should also be studied.

References

[1] S. S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Norwood, MA, USA: Artech House, 1999.

[2] P. K. Varshney, E. Masazade, P. Ray, and R. Niu, "Distributed detection in wireless sensor networks," in *Distributed Data Fusion for Network-Centric Operations*, D. Hall, C.-Y. Chong, J. Llinas, and M. Liggins, Eds. Boca Raton, FL, USA: CRC Press, 2012, ch. 4.

[3] N. Kimura and S. Latifi, "A survey on data compression in wireless sensor networks," in *Proceedings of the IEEE International Conference on Information Technology: Coding and Computing*, vol. 2, Las Vegas, NV, USA, April 2005, pp. 8–13.

[4] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82(Series D), pp. 35–45, 1960.

[5] S. J. Julier and J. K. Uhlmann, "General decentralized data fusion with covariance intersection," in *Handbook of Multisensor Data Fusion: Theory and Practice*, M. Liggins, D. Hall, and J. Llinas, Eds. Boca Raton, FL, USA: CRC Press, 2009, ch. 14.

[6] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused-track covariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 6, pp. 803–805, Nov. 1986.

[7] X. Tian and Y. Bar-Shalom, "On algorithms for asynchronous track-to-track fusion," in *Proceedings of the 13th IEEE International Conference on Information Fusion*, Edinburgh, Scotland, Jul. 2010.

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

[9] R. Forsling, A. Hansson, F. Gustafsson, Z. Sjanic, J. Löfberg, and G. Hendeby, "Conservative linear unbiased estimation under partially known covariances," *IEEE Transactions on Signal Processing*, 2022, doi:10.1109/TSP.2022.3179841.

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

[11] ——, "Communication efficient decentralized track fusion using selective information extraction," in *Proceedings of the 23rd IEEE International Conference on Information Fusion*, Virtual Conference, Jul. 2020.

[12] M. Greiff and K. Berntorp, "Optimal measurement projections with adaptive mixture Kalman filtering for GNSS positioning," in *Proceedings of the 2020 American Control Conference*, Denver, CO, USA, Jul. 2020, pp. 4435–4441.

[13] M. Greiff, A. Robertsson, and K. Berntorp, "MSE-optimal measurement dimension reduction in Gaussian filtering," in *Proceedings of the 2020 IEEE Conference on Control Technology and Applications (CCTA)*, Aug. 2020, pp. 126–133.

[14] K. Zhang, X. R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion - part VI: Sensor data compression," in *Proceedings of the 6th IEEE International Conference on Information Fusion*, Cairns, Queensland, Australia, Jul. 2003, pp. 221–228.

[15] H. Chen, K. Zhang, and X. R. Li, "Optimal data compression for multi-sensor target tracking with communication constraints," in *Proceedings of the 43rd IEEE Conference Decision and Control*, vol. 3, Atlantis, Paradise Island, Bahamas, Dec. 2004, pp. 2650–2655.

[16] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1631–1639, 2005.

[17] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4284–4299, 2007.

[18] J. Fang and H. Li, "Optimal/near-optimal dimensionality reduction for distributed estimation in homogeneous and certain inhomogeneous scenarios," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4339–4353, 2010.

[19] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1998.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[21] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice Hall, 2000.

[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 2012.

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

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

[25] B. Ghojogh, F. Karray, and M. Crowley, "Eigenvalue and generalized eigenvalue problems: Tutorial," 2019.

[26] A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2017.

[27] X. R. Li and Z. Zhao, "Measuring estimator's credibility: Noncredibility index," in *Proceedings of the 9th IEEE International Conference on Information Fusion*, Florence, Italy, Jul. 2006.

[28] J. H. Taylor, "The Cramér-Rao estimation error lower bound computation for deterministic nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 24, no. 2, pp. 343–344, April 1979.

[29] C. Fritsche, U. Orguner, and F. Gustafsson, "On parametric lower bounds for discrete-time filtering," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, Shanghai, China, Mar. 2016, pp. 4338–4342.

[30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.

[31] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.