An Optimal Data Fusion Algorithm in the Presence of Unknown Cross Covariances

Xiaohai Zhang, *Member, IEEE*

***Abstract*—This paper presents an optimal data fusion formulation and algorithm in the sense of minimum mean-square error when some or all cross covariances are unknown. This algorithm is generic in that it is capable of processing any number of measurement vectors of any dimension with any pattern of unknown cross covariances. Closed-form solution is provided for the case when all cross covariances are unknown and all measurement covariance matrices are diagonal. Numerical projected subgradient optimal fusion algorithm is provided for the most generic case. The well known covariance intersection method is shown to have a weaker upper bound of this formulation.**

***Index Terms*—Data fusion, Kalman, optimal, subgradient, unknown cross covariance.**

I. Introduction

Data fusion is the process of combining multiple measurements to refine state estimate and prediction.

When independence can be assumed among measurements, or when the cross-covariance matrices are known among measurements, the closed-form solution with linear matrix weights is available from [1]. The famous Kalman filtering [2] formulation can be viewed as a data fusion of predictions based on a system model and independent sensor measurements when the measurement matrix is orthogonal.

In real-world applications, however, cross covariances among measurements are often unknown [3], especially within a distributed fusion architecture [4], due to lack of knowledge of the true system model, double counting caused by network loops, or difficulty in acquiring reliable cross-covariance values. One of the most popular algorithms for data fusion in the presence of unknown cross covariances is the covariance intersection (CI) method which was introduced by Julier and Uhlmann [5]. While the concepts of *consistency* and *nondivergence* introduced by CI [5] play a vital role in distributed fusion architecture, the CI method only yields an overestimation of the intersection of individual covariances, which is known to be overly conservative [6].

Various efforts were made in the last two decades to achieve better estimate than the CI method in the sense of mean-square error (MSE). The largest ellipsoid (LE) method relies on the independence assumption to achieve a less conservative estimate, while the internal ellipsoidal approximation (IEA) is based on heuristics [4]. Based on Bar-Shalom Campo (BC) formula, Bakr and Lee [7] provide an optimal fusion method that is applicable for fusing only two measurements, and both measurements must be scalar or their covariance matrices must be diagonal. Based on game theory, Leonardos and Daniilidis [6] proposed the infeasible start Newton method, which provides the optimal solution for fusing two measurement vectors with unknown cross covariances.

The main contribution of this paper is an optimal data fusion formulation and algorithm in the sense of minimum MSE (MMSE) when some cross covariances are unknown, which is generic enough to deal with an arbitrary number of measurements of arbitrary dimension. Additionally, the algorithm can accommodate any mixture of known and unknown cross covariances. The relationship of our formulation with the well known CI method is also demonstrated. Application of the optimal data fusion formulation in this paper provides theoretically sound estimates that guarantee convergence and are less conservative than CI [5] in the sense of MMSE.

The paper is structured as follows. In Section II we introduce notation and formalize our problem at hand. In Section III, we present our main theoretical result, formulating an optimal upper bound of MSE in the presence of unknown cross covariances. Closed-form solution is then provided for the simplified case when all measurement covariance matrices are diagonal and all cross covariances are unknown. Finally the CI method is shown to be an upper bound of our optimal formulation. In Section IV, we present our projected subgradient optimal fusion (PSOF) method. We conclude the paper in Section V with some numerical examples and simulations.

II. Problem Formalization

In this section, we formalize the problem at hand. We start with our notation.

Let the number of measurements be N, and each measurement vector be of dimension M, where N and M are both arbitrary positive integers. We use i and j to index the measurements, and use k and l to index the dimensions. Unless explicitly specified otherwise, expressions including indices of i, j, k, and l are assumed to hold for all $i, j = 1, 2, \ldots, N$ and $k, l = 1, 2, \ldots, M$. We use bold capital letters to denote matrices, use bold small letters to denote vectors, and use normal letters to denote scalars.

Denote the N measurements for a random vector $\boldsymbol{x}$ by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$. The measurements are unbiased, i.e., $E[\boldsymbol{x}_i] = E[\boldsymbol{x}]$. Denote the covariance matrix for ith measurement by $\boldsymbol{V}_{ii}$, and the cross-covariance matrix between the ith and jth measurements by $\boldsymbol{V}_{ij}$. Then, $\boldsymbol{V}_{ij} = \boldsymbol{V}_{ji}^T$, where T is the transpose operation. Denote the joint covariance matrix by

$$\mathcal{V} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} & \cdots & \boldsymbol{V}_{1N} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} & \cdots & \boldsymbol{V}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{V}_{N1} & \boldsymbol{V}_{N2} & \cdots & \boldsymbol{V}_{NN} \end{bmatrix}. \tag{1}$$

Manuscript received September 9, 2018; revised September 13, 2018 and May 6, 2019; accepted June 21, 2019. Date of publication June 27, 2019; date of current version February 27, 2020. Recommended by Associate Editor G. Gu.

The author is with the LinkedIn Corporation, San Francisco, CA 94105 USA (e-mail: xiaohai@yahoo.com).

Digital Object Identifier 10.1109/TAC.2019.2925500

Next, we introduce a rudimental statistical result.

Lemma 1: Given random variable $\boldsymbol{x}$ with covariance matrix $\boldsymbol{V}$, the MMSE estimate of $\boldsymbol{x}$ is $\bar{\boldsymbol{x}} \equiv E[\boldsymbol{x}]$, and the resulting MSE is $Tr(\boldsymbol{V})$.

Now, consider a generic matrix-weighted linear estimator

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i$$

where $\boldsymbol{A}_i$ is any $M \times M$ real-valued matrix. For the estimator to be unbiased, we must have constraint

$$\sum_{i=1}^{N} \boldsymbol{A}_i = \boldsymbol{I} \tag{2}$$

where $\boldsymbol{I}$ is an $M \times M$ identity matrix. The covariance is

$$Cov(\hat{\boldsymbol{x}}) = \sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{V}_{ii} \boldsymbol{A}_i{}^T + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T. \tag{3}$$

To have an optimal estimator in the sense of MMSE, from Lemma 1, the problem is then to choose the values of $\boldsymbol{A}_i$ to minimize the *trace* of the covariance defined in (3) subject to the constraint in (2). When all $\boldsymbol{V}_{ii}$s and $\boldsymbol{V}_{ij}$s are known, utilizing Lagrange multipliers leads to the results in [1].

Denote the set of indices of unknown cross covariances by

$$\mathcal{D} = \{(m,n) : \boldsymbol{V}_{mn} \text{ is unknown}\}$$

and denote the set of indices of known cross covariances by

$$\mathcal{D}^c = \{(m,n) : \boldsymbol{V}_{mn} \text{ is known}\}$$

where $1 \leq m \leq N, 1 \leq n \leq N, m \neq n$.

For any $(i,j) \in \mathcal{D}$, $Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T\right)$ is not available. We can use the supremum of the trace over valid $\boldsymbol{V}_{ij}$s, i.e.,

$$\sup_{\boldsymbol{V}_{ij}} \left\{Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T\right)\right\}.$$

The problem at hand can be formulated as follows.

Problem 1: For a matrix-weighted linear estimator

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i$$

find optimal values for matrices $\boldsymbol{A}_i$ that minimizes

$$\begin{aligned} Mse(\hat{\boldsymbol{x}}) &= \sup_{\boldsymbol{V}_{ij},(i,j)\in\mathcal{D}} \{Tr(Cov(\hat{\boldsymbol{x}}))\} \\ &= \sum_{i=1}^{N} Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ii} \boldsymbol{A}_i{}^T\right) \\ &\quad + \sum_{(i,j)\in\mathcal{D}^c} Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T\right) \\ &\quad + \sum_{(i,j)\in\mathcal{D}} \sup_{\boldsymbol{V}_{ij}} \left\{Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T\right)\right\} \end{aligned} \tag{4}$$

subject to the constraint in (2).

III. Matrix-Weighted Optimal Linear Filtering

We start this section by pointing out the convexity of the optimization problem at hand. We first quote a known result.

Proposition 2: For any given set S, if $f(x,y)$ is convex in x for each $y \in S$, then $g(x) = \sup_{y\in S}\{f(x,y)\}$ is convex.

Now we present our first result in this section.

Lemma 3: The expression (4) is convex with respect to $\boldsymbol{A}_i$ for $i = 1, 2, \ldots, N$.

Proof: Since the matrix in (1) is positive semidefinite, the expression (4) before the sup operator is convex with respect to $\boldsymbol{A}_i$. Applying Proposition 2 completes the proof.

The convexity of (4) guarantees convergence when the subgradient algorithms in the next section are applied.

For a matrix $\boldsymbol{A}$ of dimension $M \times M$, denote its singular values in a nonincreasing order by $\sigma_1(\boldsymbol{A}) \geq \cdots \geq \sigma_M(\boldsymbol{A})$.

A matrix $\boldsymbol{A}$ is called a *contraction* matrix if $\sigma_1(\boldsymbol{A}) \leq 1$. The matrix contraction criterion from [5] is restated as follows.

Proposition 4: For any positive semidefinite matrix $\boldsymbol{A}$ and $\boldsymbol{B}$, the matrix

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{bmatrix}$$

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

$$\boldsymbol{X} = \boldsymbol{A}^{1/2} \boldsymbol{\Omega} \boldsymbol{B}^{1/2}.$$

From Proposition 4, the matrix $\boldsymbol{V}_{ij}$ is valid if and only if there exists a contraction matrix $\boldsymbol{\Omega}_{ij}$ such that $\boldsymbol{V}_{ij} = \boldsymbol{V}_{ii}{}^{1/2} \boldsymbol{\Omega}_{ij} \boldsymbol{V}_{jj}{}^{1/2}$. Thus, we have

$$\begin{aligned} &\sup_{\boldsymbol{V}_{ij}} \left\{Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T\right)\right\} \\ &\quad = \sup_{\boldsymbol{\Omega}_{ij}:\sigma_1(\boldsymbol{\Omega}_{ij})\leq 1} \left\{Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ii}{}^{1/2} \boldsymbol{\Omega}_{ij} \boldsymbol{V}_{jj}{}^{1/2} \boldsymbol{A}_j{}^T\right)\right\} \end{aligned} \tag{5}$$

Covariance matrices are symmetric and diagonalizable. Let $\boldsymbol{V}_{ii} = \boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{U}_i{}^T$ where $\boldsymbol{U}_i$ is an orthogonal matrix and $\boldsymbol{\Lambda}_i$ is a diagonal matrix consisting of eigenvalues of $\boldsymbol{V}_{ii}$. Then, $\boldsymbol{V}_{ii}{}^{1/2} = \boldsymbol{U}_i \boldsymbol{\Lambda}_i{}^{1/2} \boldsymbol{U}_i{}^T$. To simplify formulation, let

$$\boldsymbol{A}_i' \equiv \boldsymbol{A}_i \boldsymbol{U}_i \boldsymbol{\Lambda}_i{}^{1/2}. \tag{6}$$

Then

$$\boldsymbol{A}_i = \boldsymbol{A}_i' \boldsymbol{\Lambda}_i{}^{-1/2} \boldsymbol{U}_i{}^T. \tag{7}$$

And (5) can be written as

$$\begin{aligned} &\sup_{\boldsymbol{V}_{ij}} \left\{Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T\right)\right\} \\ &\quad = \sup_{\boldsymbol{\Omega}_{ij}:\sigma_1(\boldsymbol{\Omega}_{ij})\leq 1} \left\{Tr\left(\boldsymbol{A}_i' \boldsymbol{U}_i{}^T \boldsymbol{\Omega}_{ij} \boldsymbol{U}_j \boldsymbol{A}_j'{}^T\right)\right\} \\ &\quad = \sup_{\boldsymbol{\Omega}_{ij}:\sigma_1(\boldsymbol{\Omega}_{ij})\leq 1} \left\{Tr\left(\boldsymbol{A}_i' \boldsymbol{\Omega}_{ij} \boldsymbol{A}_j'{}^T\right)\right\}. \end{aligned} \tag{8}$$

The validity of the last step in (8) is due to the fact that a contraction matrix multiplied by orthogonal matrices has no change in its singular values and remains a contraction matrix.

Here, we introduce a couple of results from [8].

Proposition 5: For any real valued $M \times M$ matrix $\boldsymbol{A}$, we have

$$Tr(\boldsymbol{A}) \leq |Tr(\boldsymbol{A})| \leq \sum_{k=1}^{M} \sigma_k(\boldsymbol{A}).$$

Proposition 6: For any two matrices $\boldsymbol{A} \in \mathbb{C}^{m\times r}$ and $\boldsymbol{B} \in \mathbb{C}^{r\times n}$, let $q = \min\{m, r, n\}$. We have

$$\sum_{k=1}^{q} \sigma_k(\boldsymbol{AB}) \leq \sum_{k=1}^{q} \sigma_k(\boldsymbol{A}) \sigma_k(\boldsymbol{B}).$$

Next, we provide a closed-form formula for the supremum.

Corollary 7:

$$\sup_{\boldsymbol{V}_{ij}} \left\{ Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j{}^T\right)\right\} = \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_i' \boldsymbol{A}_j'{}^T\right)$$

Proof: From Propositions 5 and 6, we have

$$\begin{aligned} Tr\left(\boldsymbol{A}_i' \boldsymbol{\Omega}_{ij} \boldsymbol{A}_j'{}^T\right) &\le \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i' \boldsymbol{\Omega}_{ij}\right) \\ &\le \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i'\right) \sigma_k \left(\boldsymbol{\Omega}_{ij}\right) \\ &\le \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i'\right). \end{aligned}$$

On the other hand, let the singular value decomposition of $\boldsymbol{A}_j'{}^T \boldsymbol{A}_i'$ be $\boldsymbol{C}_{ij} \boldsymbol{\Sigma}_{ij} \boldsymbol{D}_{ij}$, where $\boldsymbol{C}_{ij}$ and $\boldsymbol{D}_{ij}$ are orthogonal matrices and $\boldsymbol{\Sigma}_{ij}$ is diagonal. Let $\boldsymbol{\Omega_{ij}} = \boldsymbol{D}_{ij}^T \boldsymbol{C}_{ij}^T$, then

$$\begin{aligned} Tr\left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i' \boldsymbol{\Omega}_{ij}\right) &= Tr\left(\boldsymbol{C}_{ij} \boldsymbol{\Sigma}_{ij} \boldsymbol{D}_{ij} \boldsymbol{D}_{ij}^T \boldsymbol{C}_{ij}^T\right) \\ &= Tr\left(\Sigma\right) = \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i'\right). \end{aligned}$$

Thus

$$\sup_{\boldsymbol{\Omega}_{ij}:\sigma_1(\boldsymbol{\Omega}_{ij})\le 1} \left\{ Tr\left(\boldsymbol{A}_i' \boldsymbol{\Omega}_{ij} \boldsymbol{A}_j'{}^T\right)\right\} = \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i'\right).$$

Combining with (8) completes the proof.

Applying Corollary 7 towards Problem 1 leads to the main results of this section.

Theorem 8: For a matrix-weighted linear estimator

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i$$

the optimal values of $\boldsymbol{A}_i$ in the sense of MMSE is given by

$$\boldsymbol{A}_i = \boldsymbol{A}_i' \boldsymbol{\Lambda}_i{}^{-1/2} \boldsymbol{U}_i{}^T$$

where $\boldsymbol{V}_{ii} = \boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{U}_i{}^T$ is the diagonalization for $\boldsymbol{V}_{ii}$, and $\boldsymbol{A}_i'$s are the values that minimizes the term

$$\begin{aligned} Mse(\hat{\boldsymbol{x}}) = &\sum_{i=1}^{N} Tr\left(\boldsymbol{A}_i' \boldsymbol{A}_i'{}^T\right) \\ &+ \sum_{(i,j)\in\mathcal{D}^c} Tr\left(\boldsymbol{A}_i' \boldsymbol{V}'_{ij} \boldsymbol{A}_j'{}^T\right) \\ &+ \sum_{(i,j)\in\mathcal{D}} \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i'\right) \end{aligned} \tag{9}$$

subject to the constraint

$$\sum_{i=1}^{N} \boldsymbol{A}_i' \boldsymbol{\Lambda}_i{}^{-1/2} \boldsymbol{U}_i{}^T = \boldsymbol{I} \tag{10}$$

where

$$\boldsymbol{V}'_{ij} \equiv \boldsymbol{\Lambda}_i{}^{-1/2} \boldsymbol{U}_i{}^T \boldsymbol{V}_{ij} \boldsymbol{U}_j \boldsymbol{\Lambda}_j{}^{-1/2}. \tag{11}$$

In the next section, we will give a numerical algorithm to compute the optimal values of $\boldsymbol{A}_i$s that minimize MSE. Once the optimal $\boldsymbol{A}_i'$ and $\boldsymbol{A}_i$ are computed, we can use the following theorem to recover the maximizing $\boldsymbol{V}_{ij}$.

Theorem 9: Let $\boldsymbol{A}_i'{}^{\text{opt}}$ be the optimal values for $\boldsymbol{A}_i'$ in the sense of MMSE. Let the singular value decomposition of $\boldsymbol{A}_j'{}^{\text{opt}T} \boldsymbol{A}_i'{}^{\text{opt}}$ be $\boldsymbol{C}_{ij} \boldsymbol{\Sigma}_{ij} \boldsymbol{D}_{ij}$, where $\boldsymbol{C}_{ij}$ and $\boldsymbol{D}_{ij}$ are orthogonal matrices and $\boldsymbol{\Sigma}_{ij}$ is diagonal. Then

$$\boldsymbol{V}_{ij}^{\text{opt}} \equiv \boldsymbol{V}_{ii}{}^{1/2} \boldsymbol{U}_i \boldsymbol{D}_{ij}^T \boldsymbol{C}_{ij}^T \boldsymbol{U}_j^T \boldsymbol{V}_{jj}{}^{1/2} \tag{12}$$

maximizes the trace item $Tr\left(\boldsymbol{A}_i \boldsymbol{V}_{ij} \boldsymbol{A}_j^T\right)$.

A. Measurements With Diagonal Covariance Matrices

In this section, we look at the special case where all measurement covariance matrices are diagonal, and all cross-covariance matrices are missing.

Having diagonal covariance matrices means that $\boldsymbol{V}_{ii} = \boldsymbol{\Lambda}_i$ and $\boldsymbol{U}_i = \boldsymbol{I}$ for $i = 1, 2, \ldots, N$. Let $\boldsymbol{\Lambda}_i = \text{diag}(\boldsymbol{\lambda}_i)$ where $\boldsymbol{\lambda}_i$ is a vector consisting of the diagonal elements of $\boldsymbol{\Lambda}_i$, and let the elements be denoted by λ_{ik} for $k = 1, 2, \ldots, M$.

The optimization problem from Theorem 8 is to minimize

$$Mse(\hat{\boldsymbol{x}}) = \sum_{i=1}^{N} Tr\left(\boldsymbol{A}_i' \boldsymbol{A}_i'{}^T\right) + \sum_{\substack{i,j=1 \\ i\ne j}}^{N} \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'{}^T \boldsymbol{A}_i'\right)$$

under the constraint of

$$\sum_{i=1}^{N} \boldsymbol{A}_i' \boldsymbol{\Lambda}_i{}^{-1/2} = \boldsymbol{I}. \tag{13}$$

Let us recall another result from [8].

Proposition 10: Let $\boldsymbol{A}$ be a matrix of dimension $m \times n$, and let $q = \min\{m, n\}$. We have

$$\sum_{k=1}^{q} |a_{ii}| \le \sum_{k=1}^{q} \sigma_k \left(\boldsymbol{A}\right).$$

Let $\boldsymbol{A}_i' = [\boldsymbol{a}_{i1}', \boldsymbol{a}_{i2}', \ldots, \boldsymbol{a}_{iM}']$ where $\boldsymbol{a}_{ik}'$ is the kth column vector of $\boldsymbol{A}_i$. Denote $\boldsymbol{a}_{ik}'$'s elements by a_{ikl}' for $l = 1, 2, \ldots, M$. Then, a_{ikl}' is $\boldsymbol{A}_i'$'s element at location (l, k). And let $\boldsymbol{1}_k$ be a vector whose element at location k is one while all other elements are zero. Denote the inner product of two vectors by $< \cdot, \cdot >$. Then, the constraint can be rewritten as

$$\sum_{i=1}^{N} \boldsymbol{a}_{ik}' \lambda_{ik}{}^{-1/2} = \boldsymbol{1}_k \qquad \text{for } k = 1, 2, \ldots, M. \tag{14}$$

Applying Proposition 10, we have

$$\begin{aligned} Mse(\hat{\boldsymbol{x}}) &\ge \sum_{i=1}^{N} Tr\left(\boldsymbol{A}_i' \boldsymbol{A}_i'{}^T\right) + \sum_{\substack{i,j=1 \\ i\ne j}}^{N} \sum_{k=1}^{M} \left|\langle\, \boldsymbol{a}_{ik}', \boldsymbol{a}_{jk}' \rangle\right| \\ &= \sum_{k=1}^{M} \left(\sum_{i=1}^{N} \langle\, \boldsymbol{a}_{ik}', \boldsymbol{a}_{ik}' \rangle + \sum_{\substack{i,j=1 \\ i\ne j}}^{N} \left|\langle\, \boldsymbol{a}_{ik}', \boldsymbol{a}_{jk}' \rangle\right| \right). \end{aligned} \tag{15}$$

Instead of directly minimizing $Mse(\hat{\boldsymbol{x}})$, let us first focus on minimizing the lower bound in (15) for a moment. Once the new minimization problem is solved, we will demonstrate that at the minimizing point, $Mse(\hat{\boldsymbol{x}})$ and the lower bound take the same value, and thus, are simultaneously minimized to the same minimum. That is, we now try to minimize

$$f(\hat{\boldsymbol{x}}) = \sum_{k=1}^{M} \left(\sum_{i=1}^{N} \langle\, \boldsymbol{a}_{ik}', \boldsymbol{a}_{ik}' \rangle + \sum_{\substack{i,j=1 \\ i\ne j}}^{N} \left|\langle\, \boldsymbol{a}_{ik}', \boldsymbol{a}_{jk}' \rangle\right| \right) \tag{16}$$

under the constraint (14).

Both the term under optimization in (16) and the constraint (14) have disjoint parameters for different k. We can optimize each kth term independently for $k = 1, 2, \ldots, M$, i.e., we are to minimize

$$f_k(\hat{\boldsymbol{x}}) = \sum_{i=1}^{N} \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{ik} \rangle + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \left| \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right|$$

$$= \sum_{i,j=1}^{N} \left| \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right| \tag{17}$$

under the constraint (14).

Since $f_k(\hat{\boldsymbol{x}})$ is not differentiable with respect to $\boldsymbol{a}'_{ik}$, we need to rely on subdifferentials. Let us define a set-valued map

$$\text{sign}(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ [-1, 1] & \text{if } x = 0 \\ \{-1\} & \text{if } x < 0 \end{cases}.$$

By [9], it is the subdifferential of $|x|$, i.e.,

$$\partial(\text{abs})(x) = \text{sign}(x)$$

where $\text{abs}(x) = |x|$.

Now, let us take subdifferential of the Lagrange term

$$\mathcal{L} = \sum_{i,j=1}^{N} \left| \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right| - \left\langle \boldsymbol{w}_k, \sum_{j=1}^{N} \lambda_{jk}^{-1/2} \boldsymbol{a}'_{jk} - \mathbf{1}_k \right\rangle$$

with respect to $\boldsymbol{a}'_{ik}$, where $\boldsymbol{w}_k$ is the Lagrange multiplier vector. According to convex function minimization rule [9], at the minimizing point, we have

$$\mathbf{0} \in \sum_{j=1}^{N} \text{sign}\left(\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right) \boldsymbol{a}'_{jk} - \frac{1}{2} \lambda_{ik}^{-1/2} \boldsymbol{w}_k \tag{18}$$

for $i = 1, 2, \ldots, N$.

Taking a dot product between $\boldsymbol{a}'_{ik}$ and both sides of the relationship (18) leads to

$$0 \in \sum_{j=1}^{N} \text{sign}\left(\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right) \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle - \frac{1}{2} \lambda_{ik}^{-1/2} \langle \boldsymbol{a}'_{ik}, \boldsymbol{w}_k \rangle.$$

Note that

$$\text{sign}\left(\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right) \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle = \left\{ \left| \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right| \right\}$$

we have

$$\sum_{j=1}^{N} \left| \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right| = \frac{1}{2} \lambda_{ik}^{-1/2} \langle \boldsymbol{a}'_{ik}, \boldsymbol{w}_k \rangle \tag{19}$$

and

$$\langle \boldsymbol{a}'_{ik}, \boldsymbol{w}_k \rangle \geq 0 \qquad \text{for } i = 1, 2, \ldots, N. \tag{20}$$

For any $l \neq i$, further take a dot product between $\boldsymbol{a}'_{lk}$ and both sides of the relationship (18), we have

$$0 \in \sum_{j=1}^{N} \text{sign}\left(\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right) \langle \boldsymbol{a}'_{lk}, \boldsymbol{a}'_{jk} \rangle - \frac{1}{2} \lambda_{ik}^{-1/2} \langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle$$

or

$$\sum_{j=1}^{N} \text{sign}\left(\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right) \langle \boldsymbol{a}'_{lk}, \boldsymbol{a}'_{jk} \rangle \ni \frac{1}{2} \lambda_{ik}^{-1/2} \langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle. \tag{21}$$

Equation (19) is true for any $1 \leq i \leq N$, we have

$$\sum_{j=1}^{N} \left| \langle \boldsymbol{a}'_{lk}, \boldsymbol{a}'_{jk} \rangle \right| = \frac{1}{2} \lambda_{lk}^{-1/2} \langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle. \tag{22}$$

From (21) and (22), we have

$$\lambda_{ik}^{-1/2} \langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle \leq \lambda_{lk}^{-1/2} \langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle. \tag{23}$$

Let

$$\lambda_k^{\min} \equiv \min_{1 \leq i \leq N} \{\lambda_{ik}\}. \tag{24}$$

Then, for any $1 \leq l \leq N$

$$(\lambda_k^{\min})^{-1/2} \langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle \leq \lambda_{lk}^{-1/2} \langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle. \tag{25}$$

Inequalities (20) and (25) dictate that, for any $1 \leq l \leq N$ with $\lambda_{lk} > \lambda_k^{\min}$, we must have $\langle \boldsymbol{a}'_{lk}, \boldsymbol{w}_k \rangle = 0$. Further more, (22) leads to $\langle \boldsymbol{a}'_{lk}, \boldsymbol{a}'_{jk} \rangle = 0$ for all $1 \leq j \leq N$.

For those $1 \leq i, l \leq N$ with $\lambda_{ik} = \lambda_{lk} = \lambda_k^{\min}$, combining relationship (21) and (22), we get

$$\sum_{j=1}^{N} \left| \langle \boldsymbol{a}'_{lk}, \boldsymbol{a}'_{jk} \rangle \right| \in \sum_{j=1}^{N} \text{sign}\left(\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \right) \langle \boldsymbol{a}'_{lk}, \boldsymbol{a}'_{jk} \rangle. \tag{26}$$

In relationship (26), every term in the left side is no less than the supremum of the corresponding right-hand side term. For any nonzero vector $\boldsymbol{a}'_{lk}$, $\langle \boldsymbol{a}'_{lk}, \boldsymbol{a}'_{lk} \rangle > 0$. Take note of the summation in the right-hand side of relationship (26) when the index j takes the value of l. The term coefficient set $\text{sign}\left(\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{lk} \rangle \right)$ must not be $\{-1\}$ for the relationship to hold. Thus, $\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{lk} \rangle \geq 0$ for any $i, l = 1, 2, \ldots, N$ with $\lambda_{ik} = \lambda_{lk} = \lambda_k^{\min}$.

By now, we can conclude that, for $k = 1, 2, \ldots, M$, when $\boldsymbol{a}'_{ik}$, $i = 1, 2, \ldots, N$, to minimize the $f_k(\hat{\boldsymbol{x}})$, we must have $\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{lk} \rangle \geq 0$ for any $i, l = 1, 2, \ldots, N$. Thus, minimization of (17) under the constraint (14) is equivalent to minimizing

$$g_k(\hat{\boldsymbol{x}}) = \sum_{i,j=1}^{N} \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle = \left\langle \sum_{i=1}^{N} \boldsymbol{a}'_{ik}, \sum_{i=1}^{N} \boldsymbol{a}'_{ik} \right\rangle \tag{27}$$

under the constraints of

$$\langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \geq 0 \qquad \text{for } i, j = 1, 2, \ldots, N \tag{28}$$

$$\sum_{i=1}^{N} \boldsymbol{a}'_{ik} \lambda_{ik}{}^{-1/2} = \mathbf{1}_k. \tag{29}$$

Our next observation is that taking the dot product between both sides of (29) with $\boldsymbol{a}'_{jk}$ leads to

$$a'_{jkk} = \sum_{i=1}^{N} \langle \boldsymbol{a}'_{ik}, \boldsymbol{a}'_{jk} \rangle \lambda_{ik}{}^{-1/2} \geq 0. \tag{30}$$

Setting $a'_{jkl} = 0$ for all $1 \leq j \leq N$, $1 \leq l \leq M, l \neq k$ leads the term under optimization in (27) to be smaller or the same, due to the fact that all vector components other than the kth of $\sum_{i=1}^{N} \boldsymbol{a}'_{ik}$ become zero, with no violation in requirements (28) and (29). So, for minimizing $\boldsymbol{a}'_{ik}$, we can safely assume $a'_{ikl} = 0$ for all $1 \leq i \leq N$, $1 \leq l \leq M, l \neq k$, i.e., elements of $\boldsymbol{a}'_{ik}$ are all zero except the kth component. The minimization problem can be further simplified as to minimize

$$h_k(\hat{\boldsymbol{x}}) = \left(\sum_{i=1}^{N} a'_{\text{ikk}} \right)^2 \tag{31}$$

under the constraint of

$$a'_{\text{ikk}} \geq 0 \qquad \text{for } i = 1, 2, \ldots, N \tag{32}$$

$$\sum_{i=1}^{N} a'_{\text{ikk}} {\lambda_{ik}}^{-1/2} = 1. \tag{33}$$

One suitable solution is

$${a'_{\text{ikk}}}^{\text{opt}} = \begin{cases} 0 & \text{for } i \neq \arg\min_j \lambda_{jk} \\ \lambda_{ik}^{1/2} & \text{for } i = \arg\min_j \lambda_{jk} \end{cases}$$

where the value $\arg\min_j \lambda_{jk}$ is a shorthand writing

$$\arg\min_j \lambda_{jk} \equiv \arg\min_{j \in \{1,2,\ldots,N\}} \lambda_{jk}.$$

Finally, a suitable solution is, for $i = 1, 2, \ldots, N$

$${a'_{\text{ikl}}}^{\text{opt}} = \begin{cases} 0 & \text{for } l \neq k, 1 \leq i \leq N \\ 0 & \text{for } l = k, i \neq \arg\min_j \lambda_{jk} \\ \lambda_{ik}^{1/2} & \text{for } l = k, i = \arg\min_j \lambda_{jk} \end{cases}. \tag{34}$$

This leads to, for $i = 1, 2, \ldots, N$

$$\left({\boldsymbol{A}_i}^{\text{opt}}\right)_{kl} = \begin{cases} 0 & \text{for } l \neq k, 1 \leq i \leq N \\ 0 & \text{for } l = k, i \neq \arg\min_j \lambda_{jk} \\ 1 & \text{for } l = k, i = \arg\min_j \lambda_{jk} \end{cases} \tag{35}$$

and the fused covariance matrix is

$$\text{diag}\left(v_1, v_2, \ldots, v_M\right) \tag{36}$$

where

$$v_k = \min_{1 \leq j \leq N} \{\lambda_{jk}\} \text{ for } k = 1, 2, \ldots, M. \tag{37}$$

The solution in (34) means that all ${\boldsymbol{A}'_i}^{\text{opt}}$ matrices are diagonal, and its kth diagonal element is not zero if and only if the kth component of the ith measurement is with the minimum variance among all the N measurements. It is trivial to verify that $Mse(\hat{\boldsymbol{x}}) = f(\hat{\boldsymbol{x}}) = \sum_{k=1}^{M} f_k(\hat{\boldsymbol{x}})$ for the choice of ${\boldsymbol{A}'_i}^{\text{opt}}$, and thus, ${\boldsymbol{A}'_i}^{\text{opt}}$ minimizes $Mse(\hat{\boldsymbol{x}})$ as well and is the optimal solution.

We conclude that the solution in (35)–(37) represents an optimal solution in the sense of MMSE when all the measurement covariance matrices are diagonal and all cross-covariance matrices are missing. The minimizing solution might not be unique, especially when multiple λ_{ik} take the same minimum value.

Remark: $\arg\min_{j \in \{1,2,\ldots,N\}} \lambda_{jk}$ is expected to return a consistent single value when multiple λ_{jk} are the same.

Remark: The result in this section is consistent with the results in [7], and it generalizes the result in [7] from two measurements to any number of measurements.

B. Relationship With CI

In this section, we demonstrate the formulation in the CI [5] as an upper bound of Theorem 8 in the sense of MMSE.

In the context of CI formulation, all cross-covariance matrices are assumed to be missing. Thus, the optimization term from Theorem 8 is

$$\sum_{i=1}^{N} Tr\left(\boldsymbol{A}'_i {\boldsymbol{A}'_i}^T\right) + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \sum_{k=1}^{M} \sigma_k\left({\boldsymbol{A}'_j}^T \boldsymbol{A}'_i\right).$$

Let $0 < \omega_i \leq 1$ for $i = 1, 2, \ldots, N$ and $\sum_{i=1}^{N} \omega_i = 1$. Then

$$Tr\left(\boldsymbol{A}'_i {\boldsymbol{A}'_i}^T\right) = \sum_{k=1}^{M} \left(\sigma_k\left(\boldsymbol{A}'_i\right)\right)^2.$$

Further applying Proposition 6 leads to an upper bound as

$$\begin{aligned} &\sum_{i=1}^{N} \sum_{k=1}^{M} \left(\sigma_k\left(\boldsymbol{A}'_i\right)\right)^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \sum_{k=1}^{M} \sigma_k\left({\boldsymbol{A}'_j}^T\right) \sigma_k\left(\boldsymbol{A}'_i\right) \\ &= \sum_{i,j=1}^{N} \sum_{k=1}^{M} \sigma_k\left(\boldsymbol{A}'_j\right) \sigma_k\left(\boldsymbol{A}'_i\right) \\ &= \sum_{i,j=1}^{N} \sum_{k=1}^{M} \left(\sqrt{\frac{\omega_i}{\omega_j}} \sigma_k\left(\boldsymbol{A}'_j\right)\right) \left(\sqrt{\frac{\omega_j}{\omega_i}} \sigma_k\left(\boldsymbol{A}'_i\right)\right) \\ &\leq \sum_{i,j=1}^{N} \sum_{k=1}^{M} \frac{1}{2} \left(\frac{\omega_i}{\omega_j} \left(\sigma_k\left(\boldsymbol{A}'_j\right)\right)^2 + \frac{\omega_j}{\omega_i} \left(\sigma_k\left(\boldsymbol{A}'_i\right)\right)^2\right) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{M} \frac{1}{\omega_i} \left(\sigma_k\left(\boldsymbol{A}'_i\right)\right)^2 \\ &= \sum_{i=1}^{N} \frac{1}{\omega_i} Tr\left(\boldsymbol{A}'_i {\boldsymbol{A}'_i}^T\right). \end{aligned}$$

Let

$$\boldsymbol{U} = \left(\sum_{i=1}^{N} \omega_i {\boldsymbol{V}_{ii}}^{-1}\right)^{-1}.$$

In the CI [5], $\boldsymbol{A}_i$ is chosen as

$$\boldsymbol{A}_i = \omega_i \boldsymbol{U} {\boldsymbol{V}_{ii}}^{-1}.$$

Thus

$$\boldsymbol{A}'_i = \omega_i \boldsymbol{U} {\boldsymbol{V}_{ii}}^{-1} \boldsymbol{U}_i {\boldsymbol{\Lambda}_i}^{1/2}$$

and the upper bound to be minimized is

$$\begin{aligned} &\sum_{i=1}^{N} \omega_i Tr\left(\boldsymbol{U} {\boldsymbol{V}_{ii}}^{-1} \boldsymbol{U}_i {\boldsymbol{\Lambda}_i}^{1/2} {\boldsymbol{\Lambda}_i}^{1/2} \boldsymbol{U}_i^T {\boldsymbol{V}_{ii}}^{-1} \boldsymbol{U}\right) \\ &= \sum_{i=1}^{N} \omega_i Tr\left(\boldsymbol{U} {\boldsymbol{V}_{ii}}^{-1} \boldsymbol{U}\right) \\ &= Tr\left(\boldsymbol{U} \left(\sum_{i=1}^{N} \omega_i {\boldsymbol{V}_{ii}}^{-1}\right) \boldsymbol{U}\right) = Tr\left(\boldsymbol{U}\right). \end{aligned}$$

This is consistent with the result in [5].

IV. Projected Subgradient Methods

The terms in (9) contain singular values which are nonsmooth functions of its entries. In this section, we start with necessary formulas for application of subgradient methods. A generic projected subgradient algorithm is then provided. Finally, different strategies of choosing subgradient step sizes are compared by simulation.

Denominator layout is used for matrix calculus.

We need to use a few results from [10]. Let

$$f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 \equiv \sum_{i=1}^{n} |x_i|, \quad \boldsymbol{x} \in \boldsymbol{R}^n \tag{38}$$

where n is a positive integer representing vector dimension. Then, $f(\boldsymbol{x})$ is an absolutely symmetric function as defined in [10]. Hence, $f \circ \sigma$ is a singular-value function. For any matrix $X \in M_{n,m}$, we have $\mathbf{1} \in \partial f(\sigma(X))$, where $\mathbf{1}$ is a vector with all components as 1 and of dimension $\min\{n, m\}$.

Proposition 11: For any matrix $\boldsymbol{X}$ of dimension $M \times M$, let its singular-value decomposition be $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T$, where $\boldsymbol{\Lambda}$ is a diagonal matrix, and $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices. Then, $\boldsymbol{U}\boldsymbol{V}^T$ is a subgradient of $\sum_{i=1}^{M} \sigma_i(\boldsymbol{X})$.

Proof: $\sum_{i=1}^{M} \sigma_i(\boldsymbol{X}) = Tr(\boldsymbol{\Lambda}) = (f \circ \sigma)(\boldsymbol{X})$. From [10, Th. 7.1], the subdifferential of the singular value function is the set

$$\{\boldsymbol{U}\text{diag}(\boldsymbol{s})\boldsymbol{V}^T : \boldsymbol{s} \in \partial f(\sigma(\boldsymbol{X})), \boldsymbol{U}\text{diag}(\sigma(\boldsymbol{X})\boldsymbol{V}^T = \boldsymbol{X}\}.$$

Let $\boldsymbol{s} = \mathbf{1} \in \partial f(\sigma(X))$, then $\text{diag}(\boldsymbol{s})$ is an identity matrix, and thus, $\boldsymbol{U}\boldsymbol{V}^T$ is a subgradient.

Proposition 12: Let

$$\boldsymbol{A}_j'^T \boldsymbol{A}_i' = \boldsymbol{C}_{ij}\boldsymbol{\Lambda}_{ij}\boldsymbol{D}_{ij}^T \tag{39}$$

be the singular-value decomposition for $\boldsymbol{A}_j'^T \boldsymbol{A}_i'$. Let

$$\boldsymbol{T}_{ji} \equiv \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'^T \boldsymbol{A}_i'\right). \tag{40}$$

Then, $\boldsymbol{T}_{ji}$ has a subgradient

$$\frac{\partial \boldsymbol{T}_{ji}}{\partial \boldsymbol{A}_i'} \equiv \boldsymbol{A}_j'\boldsymbol{C}_{ij}\boldsymbol{D}_{ij}^T \tag{41}$$

with respect to $\boldsymbol{A}_i'$, and a subgradient

$$\frac{\partial \boldsymbol{T}_{ji}}{\partial \boldsymbol{A}_j'} \equiv \boldsymbol{A}_i'\boldsymbol{D}_{ij}\boldsymbol{C}_{ij}^T \tag{42}$$

with respect to $\boldsymbol{A}_j'$.

Proof: The theorem is the result of applying subdifferential chain rule on the result of Proposition 11.

Theorem 13: With respect to $\boldsymbol{A}_i', i = 1, 2, \ldots, N$, the term in (9) of Theorem (8) has a subgradient as

$$\begin{aligned} F_{\boldsymbol{A}_i'} &= 2\boldsymbol{A}_i' + 2 \sum_{\{j:(i,j)\in\mathcal{D}^c\}} \boldsymbol{A}_j'\boldsymbol{V}_{ij}'^T \\ &\quad + 2 \sum_{\{j:(i,j)\in\mathcal{D}\}} \boldsymbol{A}_j'\boldsymbol{C}_{ij}\boldsymbol{D}_{ij}^T \end{aligned} \tag{43}$$

where $\boldsymbol{C}_{ij}$ and $\boldsymbol{D}_{ij}$ are from the singular-value decomposition of $\boldsymbol{A}_j'^T \boldsymbol{A}_i'$ as in (39).

Proof:

$$2\boldsymbol{A}_i' + 2 \sum_{\{j:(i,j)\in\mathcal{D}^c\}} \boldsymbol{A}_j'\boldsymbol{V}_{ij}'^T$$

are simply differentials of

$$\sum_{i=1}^{N} Tr\left(\boldsymbol{A}_i'\boldsymbol{A}_i'^T\right) + \sum_{(i,j)\in\mathcal{D}^c} Tr\left(\boldsymbol{A}_i'\boldsymbol{V}_{ij}'\boldsymbol{A}_j'^T\right)$$

with respect to $\boldsymbol{A}_i'$ in matrix form in denominator layout. Next, we focus our efforts on the third term of (43).

From Proposition 12

$$\sum_{(i,j)\in\mathcal{D}} \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'^T \boldsymbol{A}_i'\right)$$

has a subgradient as

$$\begin{aligned} &\sum_{j:(i,j)\in\mathcal{D}} \frac{\partial \boldsymbol{T}_{ji}}{\partial \boldsymbol{A}_i'} + \sum_{j:(j,i)\in\mathcal{D}} \frac{\partial \boldsymbol{T}_{ij}}{\partial \boldsymbol{A}_i'} \\ &= \sum_{j:(i,j)\in\mathcal{D}} \boldsymbol{A}_j'\boldsymbol{C}_{ij}\boldsymbol{D}_{ij}^T + \sum_{j:(j,i)\in\mathcal{D}} \boldsymbol{A}_j'\boldsymbol{D}_{ji}\boldsymbol{C}_{ji}^T. \end{aligned}$$

From the definition in (39)

$$\boldsymbol{A}_i'^T \boldsymbol{A}_j' = \left(\boldsymbol{A}_j'^T \boldsymbol{A}_i'\right)^T.$$

We have $\boldsymbol{C}_{ji} = \boldsymbol{D}_{ij}$ and $\boldsymbol{D}_{ji} = \boldsymbol{C}_{ij}$. Thus, the previous subgradient can be rewritten as

$$2 \sum_{j:(i,j)\in\mathcal{D}} \boldsymbol{A}_j'\boldsymbol{C}_{ij}\boldsymbol{D}_{ij}^T.$$

To simplify the notation, let

$$\boldsymbol{B}_i \equiv \boldsymbol{\Lambda}_i^{-1/2}\boldsymbol{U}_i^T \tag{44}$$

for $i = 1, 2, \ldots, N$. Then, the constraint (10) becomes

$$\sum_{i=1}^{N} \boldsymbol{A}_i'\boldsymbol{B}_i = \boldsymbol{I}. \tag{45}$$

One can easily verify that

$$\boldsymbol{A}_i'^{\text{init}} = \frac{1}{N}\boldsymbol{U}_i\boldsymbol{\Lambda}_i^{1/2} \tag{46}$$

satisfies the constraint (45).

The following theorem provides the formulation for projection.

Theorem 14: For $\boldsymbol{A}_i', i = 1, 2, \ldots, N$, its projection onto the subspace described by constraint (45) is

$$P_C(\boldsymbol{A}_i') \equiv \boldsymbol{A}_i' + \left(\boldsymbol{I}_M - \sum_{j=1}^{N} \boldsymbol{A}_j'\boldsymbol{B}_j\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{B}_i^T \tag{47}$$

where

$$\boldsymbol{\Sigma} = \sum_{j=1}^{N} \boldsymbol{B}_j^T\boldsymbol{B}_j = \sum_{j=1}^{N} \boldsymbol{V}_{jj}^{-1}. \tag{48}$$

Proof: By definition of subspace projection, $\boldsymbol{A}_i'' \equiv P_C(\boldsymbol{A}_i')$ is the solution to the minimization of $\sum_{i=1}^{N} \|\boldsymbol{A}_i'' - \boldsymbol{A}_i'\|_F^2$ subject to the constraint of $\sum_{i=1}^{N} \boldsymbol{A}_i''\boldsymbol{B}_i = \boldsymbol{I}$. One can apply Lagrange multipliers to verify the conclusion.

Here, we are ready to present our PSOF Algorithm 1.

V. Numerical Examples and Simulations

In this section, we discuss examples from [1] and [6] and compare the results. Simulations were run to evaluate different strategies of choosing step sizes for the PSOF algorithm. All codes and numerical results are available in [11].

In a distributed fusion architecture, it is often difficult to model and compute cross covariances [7]. One crude option is to assume independence among sensor results. Simulation in Fig. 1 shows that crude assumption of independence leads to rapid system divergence, while PSOF applied in distributed filtering architecture can ensure system convergence with less conservative variances than the CI method.

Next, we consider two zero-mean measurements with covariance matrices

$$\boldsymbol{V}_{11} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} \quad \boldsymbol{V}_{22} = \begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}.$$

Algorithm 1: Projected Subgradient Optimal Fusion.

In : N, M the measurements count & dimension
$\mathcal{D}$ a set of indices with unknown $\boldsymbol{V}_{ij}$
$\boldsymbol{V}_{ij}$ for all $(i,j) \notin \mathcal{D}$
Out: mse estimated mean square error
$\boldsymbol{A}_i$ optimal matrix weights
$\boldsymbol{V}_{ij}^{opt}$ maximizing matrix for $(i,j) \in \mathcal{D}$
$\boldsymbol{V}_{opt}$ the output covariance

for *i=1 to N* **do**
  Let $\boldsymbol{V}_{ii} = \boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{U}_i^T$ be $\boldsymbol{V}_{ii}$'s diagonalization;
  Let $\boldsymbol{B}_i = \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{U}_i^T$;
end
Use (11) to compute $\boldsymbol{V}'_{ij}$ if $(i,j) \in \mathcal{D}$;
Let $\boldsymbol{\Sigma}^{-1} = \left(\sum_{i=1}^{N} \boldsymbol{V}_{ii}^{-1}\right)^{-1}$;
Initialize $\boldsymbol{A}'_i = \frac{1}{N} \boldsymbol{U}_i \boldsymbol{\Lambda}_i^{1/2}$ for $1 \le i \le N$;
Set $best\_mse$ to a large value;
$best\_\boldsymbol{A}'_i = \boldsymbol{A}'_i$;
repeat
  Let $\boldsymbol{A}'^{T}_j \boldsymbol{A}'_i = \boldsymbol{C}_{ij} \boldsymbol{\Lambda}_{ij} \boldsymbol{D}_{ij}^T$ be the singular value decomposition if $(i,j) \in \mathcal{D}$;
  for *i=1 to N* **do**
    Use (43) to compute the subgradient $\boldsymbol{G}_i$;
  end
  Compute a $step$ size with a chosen strategy;
  Set $\boldsymbol{A}'_i = \boldsymbol{A}'_i - \boldsymbol{G}_i \times step$;
  Use (47) to set $\boldsymbol{A}'_i = P_C(\boldsymbol{A}'_i)$;
  Compute the new mse use (9);
  if $mse < best\_mse$ **then**
    $best\_mse = mse$; $best\_\boldsymbol{A}'_i = \boldsymbol{A}'_i$;
  end
until *user defined stop condition or max iteration*;
Use (7) to compute $\boldsymbol{A}_i$ from $best\_\boldsymbol{A}'_i$;
Use (12) to compute $\boldsymbol{V}_{ij}^{opt}$ for $(i,j) \in \mathcal{D}$;
Use (3) to compute the output covariance $\boldsymbol{V}_{opt}$;

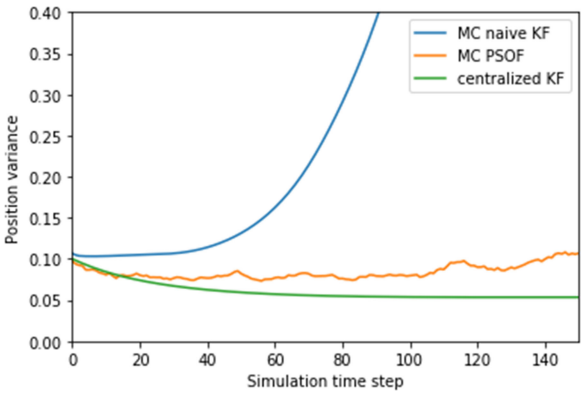

Fig. 1. Simulation of a fully distributed three node fusion architecture with parameters and structure from [1]. In the legend, Monte Carlo simulation is denoted by MC and Kalman filtering is denoted by KF. Kalman filtering with naive assumption of independence diverges quickly, while distributed fusion with PSOF converges well. Centralized Kalman filtering results are included for referencing.

The cross covariance between them is assumed to be unknown.

From (35) and (36), the optimal coefficient matrices are

$$\boldsymbol{A}_1^{\text{opt}} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{A}_2^{\text{opt}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

and the fused covariance is

$$\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}.$$

With a constant step size of $4.0e$-4, the PSOF simulation with 8000 iterations converges to

$$\boldsymbol{A}_1^{\text{opt}} = \begin{bmatrix} 1.98e\text{-}04 & 0 \\ 0 & 1.00 \end{bmatrix} \quad \boldsymbol{A}_2^{\text{opt}} = \begin{bmatrix} 1.00 & 0 \\ 0 & 3.06e\text{-}04 \end{bmatrix}.$$

The fused covariance from the simulation is

$$\begin{bmatrix} 3.00 & 0.00 \\ 0.00 & 5.00 \end{bmatrix}.$$

Both the theoretical result and the simulation are consistent with the result from [6].

For the second example from [6], we used large covariance to simulate the missing measurement component. Simulation shows similar consistency.

Finally, we compare different strategies of choosing step sizes for the PSOF algorithm. Assuming three measurements, the covariance matrices are chosen to be

$$\boldsymbol{V}_{11} = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{V}_{22} = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} \quad \boldsymbol{V}_{33} = \begin{bmatrix} 4 & 0 \\ 0 & 100 \end{bmatrix}$$

and we assume all cross-covariance matrices to be missing.

From (35) and (36), the optimal coefficient matrices are

$$\boldsymbol{A}_1^{\text{opt}} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{A}_2^{\text{opt}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \boldsymbol{A}_3^{\text{opt}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and the fused covariance is

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

The simulation results are consistent with the theoretical results.

The strategies we considered include (refer to [12]):

CSS	constant step size;
CSL	constant step length;
SSNS	square summable but not summable;
NSDSS	nonsummable diminishing step size;
NSDSL	nonsummable diminishing step length;
Polyak	Polyak's step length.

Convergence speeds of subgradient algorithms are very sensitive to the step sizes and hyper-parameters chosen. It is hard to have an impartial and reliable comparison between different strategies. The approach in this paper is to adjust the initial step sizes in a way that their stable-state MSE variances are approximately the same. With that, the convergence comparison is illustrated in Fig. 2.

Broadly speaking, SSNS, NSDSS, NSDSL, and Polyak's strategies are in the same category that converges faster than the other category consisting of CSS and CSL. While SSNS initially converges the fastest, it converges much slower in the later stage. NSDSL, CSL, and Polyak's methods incur extra computational cost of computing the norm of the $\boldsymbol{A}'_{ij}$ as evident in Fig. 3 (the absolute values matter less than the relative values). NSDSS performed reliably with no need for computing the norm of the variables.

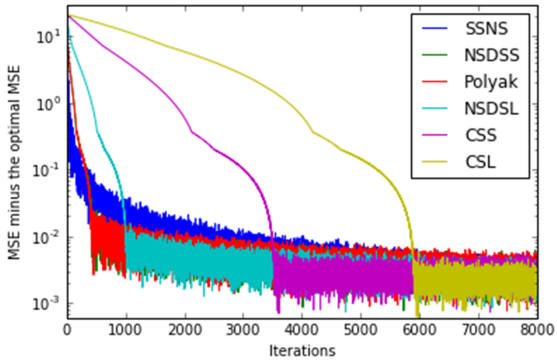

Fig. 2. Convergence speed comparison.

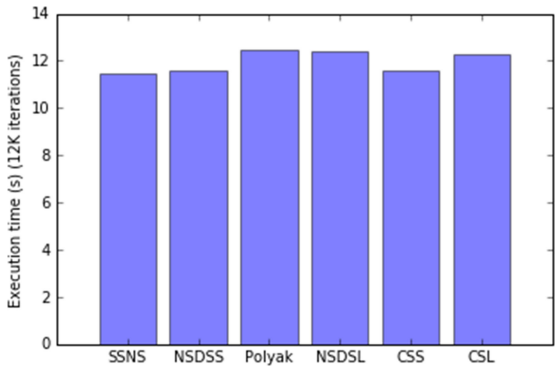

Fig. 3. Simulation execution time comparison.

Even though we only presented the basic subgradient algorithms in this paper, more advanced convex optimization algorithms can be applied to accelerate the computation.

REFERENCES

[1] S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion Kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, Jun. 2004. [Online]. Available: http://dx.doi.org/10.1016/j.automatica.2004.01.014

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.

[3] 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. [Online]. Available: http://dx.doi.org/10.1155/2015/683701

[4] M. A. Bakr and S. Lee, "Distributed multisensor data fusion under unknown correlation and data inconsistency," *Sensors*, vol. 17, no. 11, 2017, Art. no. 2472. [Online]. Available: http://www.mdpi.com/1424-8220/17/11/2472

[5] S. J. Julier and J. K. Uhlmann, "Non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. 16th Amer. Control Conf.*, Jun. 1997, vol. 4, pp. 2369–2373.

[6] S. Leonardos and K. Daniilidis, "A game-theoretic approach to robust fusion and Kalman filtering under unknown correlations," in *Proc. Amer. Control Conf.*, May 2017, pp. 2568–2573.

[7] M. A. Bakr and S. Lee, "Track level fusion with an estimation of maximum bound of unknown correlation," in *Proc. Int. Conf. Control, Autom. Inf. Sci.*, Oct. 2016, pp. 32–36.

[8] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York, NY, USA: Cambridge Univ. Press, 1986.

[9] Q. Ansari, C. S. Lalitha, and M. Mehta, Generalized Convexity, Nonsmooth Variational Inequalities, and Nonsmooth Optimization. *Boca Raton*, FL, USA: CRC Press, Jan. 2014.

[10] A. S. Lewis and H. S. Sendov, "Nonsmooth analysis of singular values. Part I: Theory," *Set-Valued Anal.*, vol. 13, no. 3, pp. 213–241, Sep. 2005. [Online]. Available: https://doi.org/10.1007/s11228-004-7197-7

[11] X. Zhang, "Numerical results for PSOF," 2018. [Online]. Available: https://github.com/xiaohai2016/psof/

[12] S. Boyd and A. Mutapcic, "Subgradient methods," 2007. [Online]. Available: https://web.stanford.edu/class/ee392o/subgrad_method.pdf