Optimal linear MMSE estimation under correlation uncertainty in restricted Bayesian framework

Berkan Dulek, Suleyman Taylan Topaloglu, and Sinan Gezici

Abstract

A restricted Bayes approach is proposed for linear estimation of a scalar random parameter based on a scalar observation under uncertainty regarding the correlation between the parameter and the observation. In particular, the optimal linear estimator that minimizes the average mean-squared error (MSE) is derived under a constraint on the worst-case MSE by considering possible values of the correlation coefficient and its probability distribution. A closed-form expression is derived for the optimal linear estimator in the proposed restricted Bayesian framework by considering a generic statistical characterization of the correlation coefficient. Performance of the proposed estimator is evaluated via numerical examples and its benefits are illustrated in various scenarios. The proposed framework is also extended to the case of vector-valued observation and the properties of the optimal linear estimator are characterized.

Index Terms

Correlation uncertainty, linear estimation, restricted Bayes, signal processing.

I. INTRODUCTION

OPTIMAL linear minimum mean-squared error (LMMSE) estimation of a random parameter based on a random (possibly vector-valued) observation requires perfect knowledge of the covariance between the desired parameter and the observation [1, Section V.C]. With this information, it becomes possible to linearly combine the observation(s) and the mean value of the desired parameter so that the resulting estimator attains the lowest mean-squared error (MSE) among all linear estimators. When the true value of the covariance is not available but only known partially, conservative estimation techniques can be

B. Dulek is with the Department of Electrical and Electronics Engineering, Hacettepe University, Beytepe Campus, Ankara 06800, Turkey. Suleyman Taylan Topaloglu and Sinan Gezici are with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey. e-mails: berkan@ee.hacettepe.edu.tr, {taylan,gezici}@ee.bilkent.edu.tr. S. T. Topaloglu is also with ASELSAN Inc., Ankara 06200, Turkey (e-mail: sttopaloglu@aselsan.com.tr).

adopted to provide performance guarantees on the accuracy of the resulting estimators over all possible values that the covariance can take.

The problem of covariance uncertainty, or equivalently correlation uncertainty in the case of random quantities with known means, has been investigated in the distributed estimation literature. In this setting, sensor nodes compute estimates of the same parameters, which are then fused into an overall estimate in order to attain improved estimation performance. To this end, the fusion of the state estimates of a target based on measurements obtained by two different sensors is considered in [2] and [3] assuming joint normal distribution for state and measurement variables. Since the same process noise in the kinematic model of the target affects both measurements, the errors in the state estimates are correlated. As a result, the optimal (in maximum-likelihood sense) fusion of two estimates requires perfect knowledge of the cross-correlation between state estimation errors, which culminate in the well-known Bar-Shalom/Campo (BC) formula. The optimal BC estimator linearly combines the local state estimates with weights computed using the auto- and cross-covariances of the local estimators.

In order to prevent erroneous estimation results due to under-assessment of the variation of the observed values, Ellipsoidal (such as Covariance Intersection, Largest Ellipsoid) and model-based techniques have been employed for the problem of fusion under unknown correlation. (See [4] and [5], and references therein.) In Ellipsoidal techniques, a covariance bound that overestimates the true covariance of the fused estimate is provided in order to perform consistent estimation without using cross-covariance matrices [6]. Since the unknown cross-correlation is modeled as a random quantity in our paper, we focus on the model-based techniques in the relevant literature (mostly within the framework of decentralized estimation). To deal with correlation uncertainty, techniques that rely on properties of the correlation or application-specific prior knowledge have been developed. For the case of scalar state estimates with two sensors, the optimal BC estimator is expressed as a function of the correlation coefficient between the local estimates in [7]. Closed-form expressions are presented for the interval of possible means and variances of the optimal estimator by using the fact that the correlation coefficient takes values in the interval $[-1, 1]$. Furthermore, uncertainty in the correlation coefficient is modeled with a uniform distribution and the expressions for the mean and the variance are provided by marginalizing out the correlation coefficient.

In scalar case, it follows from the definition of the correlation coefficient that the cross-covariance is equal to the correlation coefficient multiplied with the product of the standard deviations of the corresponding random variables as in $\mathrm{Cov}(X, Y) = \rho\,\sigma_X \sigma_Y$. This fact is employed in a number of studies to parameterize the variance of the BC estimator in terms of the unknown correlation coefficient. A generalization of the correlation coefficient to high dimensions is also presented in [8] using a Cholesky type decomposition of the unknown cross-covariance matrix in terms of the single sensor covariance matrices. This cross-covariance matrix model, parameterized with the analogously defined ‘correlation coefficient’, is employed in [9] to derive a conservative fusion rule which minimizes the maximal Mahalanobis distance with respect to the BC estimator (expressed in terms of the correlation coefficient) over all values of the correlation coefficient assuming that it lies in a subinterval of $[-1, 1]$. In addition, a closed form estimation fusion rule is also proposed by assuming a uniform prior for the correlation coefficient and computing the expectation of the BC estimator.

Uncertainty sets for the error covariance matrix of local state estimates are employed in other studies, based on which robust fusion algorithms that minimize the worst-case fused MSE are proposed. Among these are the uncertainty set constructed with norm-bounded additive perturbations in [10], that defined in [11] by placing an upper bound, termed allowance of cross-covariance, on the maximum eigenvalue of the normalized cross-covariance (NCC) matrix, and the ones defined in [12] which impose element-wise constraints on the NCC matrix.

The main conceptual difference between the decentralized estimation framework described above and the one considered in our paper is that the former considers the problem of fusing correlated local state estimates under lack of knowledge about their cross-correlation which precludes the use of the optimal BC fusion rule. Local estimates are basically observations of the fusion center. On the other hand, in our work, we consider the LMMSE estimation of a desired random parameter based on an observation of another random quantity. There is uncertainty regarding the correlation coefficient between the desired parameter and the observation, and the covariance of the observation is assumed to be known. The uncertainty in the correlation coefficient is modeled with a probability density function and we invoke results from copula theory to restrict its support instead of using the standard $[-1, 1]$ range.

As for the technical contribution, unlike the purely Bayesian and min-max based worst-case approaches employed in the estimation fusion literature, we employ a restricted Bayes approach to obtain the optimal linear estimator that minimizes the average MSE with respect to the assumed distribution while satisfying a given constraint on the worst-case MSE for all possible values of the correlation coefficient [13]. This allows us to strike any desired balance between the average and the worst-case estimation performances. Necessary and sufficient conditions for the feasibility of the proposed optimization problem are stated. The solution for the optimal linear estimator is expressed in closed-form in terms of the minimum, maximum and average values of the attainable correlation coefficients and the constraint on the worst-case MSE. The theoretical results are corroborated with numerical examples and also extended to the case of vector-valued observation.

II. PRELIMINARIES

Let X and Y be two scalar random variables with unknown joint cumulative distribution function (CDF) $F_{X,Y}(x,y)$ and known marginals $F_X(x)$ and $F_Y(y)$, respectively. The means and the variances of X and Y can be computed from the known marginal CDFs and are denoted as (μ_X, σ_X^2) and (μ_Y, σ_Y^2), respectively. The variances are assumed to be non-zero and finite. The corresponding standard deviations are denoted as σ_X and σ_Y. It is assumed that the correlation coefficient between X and Y, denoted with ρ, is unknown. It is a well-known fact that the correlation coefficient takes values in the interval $[-1, 1]$, and the extreme points -1 and 1 are achieved if and only if there is an almost sure linear relationship between X and Y.

The bounds on the correlation coefficient can be specified in a tighter form based on the attainable correlations theorem [14, Theorem 5.25]. Let $\rho_{\min}$ and $\rho_{\max}$ denote, respectively, the minimum and the maximum values that can be attained by the correlation coefficient. Then, the theorem states that the minimum correlation $\rho_{\min}$ is attained if and only if X and Y are countermonotonic and the maximum correlation $\rho_{\max}$ is attained if and only if X and Y are comonotonic. Furthermore, the values of $\rho_{\min}$ and $\rho_{\max}$ are explicitly specified based on the marginal CDFs. To this end, we first recall the Fréchet bounds for bivariate random variables [15]:

$$\max\{F_X(x) + F_Y(y) - 1, 0\} \leq F_{X,Y}(x,y) \leq \min\{F_X(x), F_Y(y)\}, \tag{1}$$

which holds for all $x \in \mathbb{R}$ and $y \in \mathbb{R}$. It is noted that the upper bound is attained when X and Y are comonotonic, i.e., $X = F_X^{-1}(U)$ and $Y = F_Y^{-1}(U)$, where U is a uniform random variable in the interval $(0,1)$, i.e., $U \sim \mathcal{U}(0,1)$. On the other hand, the lower bound is attained when X and Y are countermonotonic, i.e., $X = F_X^{-1}(U)$ and $Y = F_Y^{-1}(1-U)$ for some $U \sim \mathcal{U}(0,1)$. Applying the above bounds in the Hoeffding's covariance identity given below [14, Lemma 5.24]

$$\mathrm{Cov}(X,Y) = \int_{\mathbb{R}}\int_{\mathbb{R}} \left(F_{X,Y}(x,y) - F_X(x)F_Y(y)\right) dx dy \tag{2}$$

yields the following tight lower and upper bounds on the covariance of jointly distributed X and Y:

$$\mathrm{Cov}\left(\underline{X}, \underline{Y}\right) \leq \mathrm{Cov}\left(X, Y\right) \leq \mathrm{Cov}\left(\overline{X}, \overline{Y}\right), \tag{3}$$

where $(\overline{X}, \overline{Y}) \triangleq (F_X^{-1}(U), F_Y^{-1}(U))$ and $(\underline{X}, \underline{Y}) \triangleq (F_X^{-1}(U), F_Y^{-1}(1-U))$ denote, respectively, comonotonic and countermonotonic random variables that have the same marginals as (X, Y) (hence, the same means and variances). Normalizing with the product of the standard deviations of X and Y, we get the minimum and the maximum values for the attainable correlations:

$$\rho_{\min} \triangleq \rho(\underline{X}, \underline{Y}) \leq \rho \leq \rho(\overline{X}, \overline{Y}) \triangleq \rho_{\max} \,. \tag{4}$$

More explicitly,

$$\rho_{\max} = \frac{\mathbb{E}\left[\overline{XY}\right] - \mu_X \mu_Y}{\sqrt{\sigma_X^2 \sigma_Y^2}} \,,\ \rho_{\min} = \frac{\mathbb{E}\left[\underline{XY}\right] - \mu_X \mu_Y}{\sqrt{\sigma_X^2 \sigma_Y^2}} \,, \tag{5}$$

where $\mathbb{E}\left[\overline{XY}\right] = \int_0^1 F_X^{-1}(u) F_Y^{-1}(u) du$ and $\mathbb{E}\left[\underline{XY}\right] = \int_0^1 F_X^{-1}(u) F_Y^{-1}(1-u) du$.

We also mention the following properties of attainable correlations [14, Theorem 5.25]: (a) $\rho_{\min} < 0 < \rho_{\max}$; (b) $\rho_{\max} = 1$ if and only if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that X and $aY + b$ have the same marginal CDFs; likewise, $\rho_{\min} = -1$ holds for the same argument with $a < 0$; (c) if at least one of two random variables has a symmetric probability density/mass function with respect to its mean, we have $\rho_{\min} = -\rho_{\max}$.

III. LINEAR UNBIASED ESTIMATION UNDER CORRELATION UNCERTAINTY

We consider the problem of linear unbiased estimation of a scalar random variable X based on an observation of another scalar random variable Y. The well-known optimal linear estimator that minimizes the Bayesian MSE is given by [1]

$$\hat{X}(Y) = \mu_X + \rho \frac{\sigma_X}{\sigma_Y} (Y - \mu_Y) \,. \tag{6}$$

As seen from (6), the optimal estimator requires the exact knowledge of the correlation coefficient ρ as well as the means and the variances of X and Y.

Under uncertainty on the true value of the correlation coefficient, the design of an optimal linear unbiased estimator with respect to the restricted Bayes criterion is the subject of this paper. More explicitly, we consider the set of all linear unbiased estimators, i.e., estimators that can be expressed in the following form:

$$\hat{X}_\alpha(Y) = \mu_X + \alpha (Y - \mu_Y) \,. \tag{7}$$

The corresponding MSE is obtained as

$$\begin{aligned}\mathbb{E}\left[(X - \hat{X}_\alpha(Y))^2\right] &= \mathbb{E}\left[(X - (\mu_X + \alpha(Y - \mu_Y)))^2\right] \\ &= \sigma_X^2 - 2\alpha\,\mathrm{Cov}(X,Y) + \alpha^2\sigma_Y^2 \\ &= \sigma_X^2 - 2\alpha\rho\sigma_X\sigma_Y + \alpha^2\sigma_Y^2 \triangleq R_\rho(\alpha)\,. \end{aligned} \tag{8}$$

In the case of the Bayesian MMSE criterion, the optimal choice of α (i.e., that minimizes $R_\rho(\alpha)$) is $\alpha^* = \rho\sigma_X/\sigma_Y$, as seen from (6). It is noted that the MSE in (8), denoted with $R_\rho(\alpha)$, depends on the unknown value of the correlation coefficient ρ. We assume that the uncertainty in ρ can be modeled in the form of a probability density function (PDF) $f(\rho)$ with support in the interval $[\rho_{\min}, \rho_{\max}]$, which can be obtained from the attainable correlations theorem as explained in Section II or based on prior experience. A uniform distribution based on Laplace's principle of insufficient reason or an uninformative prior like Jeffreys' $f(\rho) = c/(1-\rho^2)$ for $\varrho \in [\rho_{\min}, \rho_{\max}]$ can be assumed for $f(\rho)$. The latter has connections with the PDF of the sample correlation coefficient for jointly distributed Gaussian random variables [16, Section 3.2]. Under the restricted Bayesian framework, we seek for the optimal linear unbiased estimator that minimizes the average MSE (averaged over the distribution of the correlation coefficient) subject to a constraint on the worst-case MSE as expressed by the following optimization problem:

$$\begin{aligned}&\min_{\alpha\in\mathbb{R}} \qquad \int_{\rho_{\min}}^{\rho_{\max}} R_\rho(\alpha) f(\rho) d\rho \\ &\text{subject to} \quad R_\rho(\alpha) \le r \qquad \forall \rho \in [\rho_{\min}, \rho_{\max}]\,,\end{aligned} \tag{P1}$$

where r specifies an upper bound on the maximum MSE of the estimator that is subject to design. In order to maintain generality, we assume $-1 \le \rho_{\min} < \rho_{\max} \le 1$. In other words, we do not restrict $\rho_{\min}$ to be negative and $\rho_{\max}$ to be positive as dictated by the attainable correlations theorem. Prior information may be available to localize ρ more accurately. The following lemma specifies the conditions for the feasibility of the optimization problem (P1).

Lemma: *The optimization problem* (P1) *is feasible if and only if the constraint on the worst MSE satisfies*

Case 1: $r \ge \sigma_X^2$ *if* $\rho_{\min} < 0 < \rho_{\max}$,

Case 2: $r \ge \sigma_X^2(1-\rho_{\min}^2)$ *if* $0 \le \rho_{\min} < \rho_{\max}$,

Case 3: $r \ge \sigma_X^2(1-\rho_{\max}^2)$ *if* $\rho_{\min} < \rho_{\max} \le 0$,

Proof: The constraint in (P1) can equivalently be written as

$$\max_{\rho\in[\rho_{\min},\rho_{\max}]} R_\rho(\alpha) \le r\,. \tag{9}$$

For the optimization problem (P1) to be feasible, the constraint parameter r should be selected as

$$r \ge \min_{\alpha\in\mathbb{R}} \max_{\rho\in[\rho_{\min},\rho_{\max}]} R_\rho(\alpha) \triangleq \min_{\alpha\in\mathbb{R}} R_{\rho^*(\alpha)}(\alpha) \tag{10}$$

Let $\rho^*(\alpha)$ denote the value of the correlation coefficient that maximizes $R_\rho(\alpha)$ for a given value of $\alpha \in \mathbb{R}$. Since $R_\rho(\alpha)$ is linear in ρ for fixed α, its maximum value occurs at an extreme point of the interval $[\rho_{\min}, \rho_{\max}]$. More explicitly, $\rho^*(\alpha) = \rho_{\min}$ for $\alpha > 0$; $\rho^*(\alpha) = \rho_{\max}$ for $\alpha < 0$; and $R_\rho(\alpha) = \sigma_X^2$ is independent of ρ if $\alpha = 0$. Hence, for $\alpha \ge 0$, we get $R_{\rho^*(\alpha)}(\alpha) = R_{\rho_{\min}}(\alpha) = \sigma_X^2 - 2\alpha\rho_{\min}\sigma_X\sigma_Y + \alpha^2\sigma_Y^2$. If $\rho_{\min} \le 0$, the minimum occurs at $\alpha = 0$ with the value of the function equal to σ_X^2. If $\rho_{\min} > 0$, the minimum occurs at $\alpha = \rho_{\min}\sigma_X/\sigma_Y$ with the value of the function equal to $\sigma_X^2(1-\rho_{\min}^2)$. On the other hand, for $\alpha \le 0$, we get $R_{\rho^*(\alpha)}(\alpha) = R_{\rho_{\max}}(\alpha) = \sigma_X^2 - 2\alpha\rho_{\max}\sigma_X\sigma_Y + \alpha^2\sigma_Y^2$. If $\rho_{\max} \ge 0$, the minimum occurs again at $\alpha = 0$ with the value of the function equal to σ_X^2. If $\rho_{\max} < 0$, the minimum occurs at $\alpha = \rho_{\max}\sigma_X/\sigma_Y$ with the value of the function equal to $\sigma_X^2(1-\rho_{\max}^2)$. Taking the minimum over both cases ($\alpha \ge 0$ and $\alpha \le 0$) and classifying the conditions based on the signs of $\rho_{\min}$ and $\rho_{\max}$, we obtain the results for the inequality in (10) as specified under the three cases in the lemma. □

Before we proceed with the solution of the optimization problem (P1), a number of remarks are in order. First, since $R_\rho(\alpha)$ is linear in the correlation coefficient ρ, its average value over the distribution of the correlation coefficient is equal to its value evaluated at the mean value of the correlation coefficient. More specifically,

$$\begin{aligned}
\mathbb{E}\left[R_\rho(\alpha)\right] &= \int_{\rho_{\min}}^{\rho_{\max}} R_\rho(\alpha) f(\rho) d\rho \\
&= \int_{\rho_{\min}}^{\rho_{\max}} \left(\sigma_X^2 - 2\alpha\rho\sigma_X\sigma_Y + \alpha^2\sigma_Y^2\right) f(\rho) d\rho \\
&= \sigma_X^2 - 2\alpha\overline{\rho}\sigma_X\sigma_Y + \alpha^2\sigma_Y^2 = R_{\overline{\rho}}(\alpha)\,,
\end{aligned} \tag{11}$$

where $\overline{\rho}$ denotes the mean value of the correlation coefficient, i.e., $\mathbb{E}[\rho] = \int_{\rho_{\min}}^{\rho_{\max}} \rho f(\rho) d\rho = \overline{\rho}$. Next, to simplify the notation, we employ a change of optimization variable by defining $\alpha = \varrho\sigma_X/\sigma_Y$. Under this parameterization, the objective function can be expressed as

$$R_{\overline{\rho}}(\alpha) = R_{\overline{\rho}}(\varrho) = \sigma_X^2(1 - 2\overline{\rho}\varrho + \varrho^2)\,, \tag{12}$$

and the constraint function can compactly be written as

$$\max_{\rho\in[\rho_{\min},\rho_{\max}]} R_\rho(\alpha) = \max_{\rho\in[\rho_{\min},\rho_{\max}]} R_\rho(\varrho) = \begin{cases} \sigma_X^2(1-2\rho_{\min}\varrho+\varrho^2) & \text{if } \varrho \geq 0 \\ \sigma_X^2(1-2\rho_{\max}\varrho+\varrho^2) & \text{if } \varrho \leq 0 \end{cases}. \tag{13}$$

Then, after proper normalization, the optimization problem (P1) can equivalently be expressed as

$$\begin{aligned} \min_{\varrho\in\mathbb{R}} \quad & \varrho^2 - 2\overline{\rho}\varrho + 1 \\ \text{subject to} \quad & \varrho^2 - 2\rho_{\min}\varrho + 1 \leq r/\sigma_X^2 \text{ if } \varrho \geq 0 \\ & \varrho^2 - 2\rho_{\max}\varrho + 1 \leq r/\sigma_X^2 \text{ if } \varrho \leq 0\,. \end{aligned} \tag{P2}$$

The following proposition presents the solution of (P2) (and hence, (P1)), i.e., the optimal linear unbiased estimator that minimizes the average MSE under the condition that the worst-case MSE does not exceed a certain value.

Proposition 1: *Let $\rho_{\max}^{(1)}$ and $\rho_{\min}^{(2)}$ be defined as follows:*

$$\begin{aligned} \rho_{\max}^{(1)} &= \rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}, \\ \rho_{\min}^{(2)} &= \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}\,. \end{aligned} \tag{14}$$

Then, the solution of (P2) *(equivalently,* (P1)*) is given as*

Case a. *If $r \geq \sigma_X^2$,*

$$\alpha^{\text{opt}} = \begin{cases} \overline{\rho}\sigma_X/\sigma_Y & \textit{if } \overline{\rho} \in [\rho_{\max}^{(1)}, \rho_{\min}^{(2)}] \\ \rho_{\min}^{(2)}\sigma_X/\sigma_Y & \textit{if } \overline{\rho} > \rho_{\min}^{(2)} \\ \rho_{\max}^{(1)}\sigma_X/\sigma_Y & \textit{if } \overline{\rho} < \rho_{\max}^{(1)} \end{cases} \tag{15}$$

Case b. *If $\sigma_X^2(1-\rho_{\min}^2) \leq r \leq \sigma_X^2$ and $0 \leq \rho_{\min} < \rho_{\max}$,*

$$\alpha^{\text{opt}} = \begin{cases} \overline{\rho}\sigma_X/\sigma_Y & \textit{if } \overline{\rho} \in [\rho_{\min}, \rho_{\min}^{(2)}] \\ \rho_{\min}^{(2)}\sigma_X/\sigma_Y & \textit{if } \overline{\rho} > \rho_{\min}^{(2)} \end{cases} \tag{16}$$

Case c. *If $\sigma_X^2(1-\rho_{\max}^2) \leq r \leq \sigma_X^2$ and $\rho_{\min} < \rho_{\max} \leq 0$,*

$$\alpha^{\text{opt}} = \begin{cases} \overline{\rho}\sigma_X/\sigma_Y & \textit{if } \overline{\rho} \in [\rho_{\max}^{(1)}, \rho_{\max}] \\ \rho_{\max}^{(1)}\sigma_X/\sigma_Y & \textit{if } \overline{\rho} < \rho_{\max}^{(1)} \end{cases} \tag{17}$$

If none of the above cases hold, then the optimization problem given in (P2) *(or,* (P1)*) is not feasible.*

Proof: Please see Appendix.

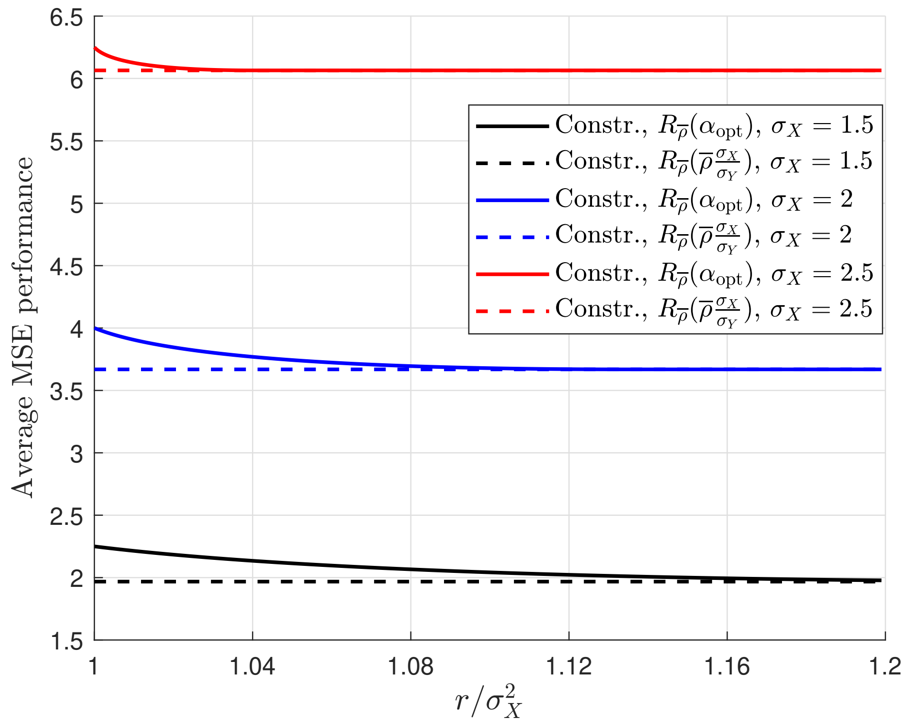

Fig. 1. The value of the objective function evaluated at the solution of the constrained optimization problem (P1) for the lognormal setting as a function of the constraint parameter r and different values of σ_X.

As a final remark in this section, we note that if the solution of the unconstrained optimization problem satisfies the constraint, it is also optimal for the constrained optimization problem given in (P1), i.e., $\alpha^{\rm opt} = \overline{\rho}\sigma_X/\sigma_Y$. If it does not satisfy the constraint on the worst-case MSE under correlation uncertainty, then the solution of the proposed optimization problem acts as a shrinkage operator. More specifically, the magnitude of linear estimation coefficient α is decreased towards zero, which effectively results in a regression towards the mean μ_X as can be seen from (7).

IV. NUMERICAL EXAMPLES

In this section, we present numerical examples for the constrained optimization problem given in (P1). In [14, Example 5.26], minimal and maximal correlations are obtained for two lognormal random variables, where $\ln X \sim \mathcal{N}(0, \sigma_X^2)$ and $\ln Y \sim \mathcal{N}(0, 1)$. Since $\ln X \sim \sigma_X \ln Y$, the condition (b) mentioned at the end of Section II is not satisfied (unless $\sigma_X \neq 1$), and hence, it holds that $\rho_{\min} > -1$ and $\rho_{\max} < 1$. In particular, it is shown in [14, Example 5.26] that the minimal and the maximal correlations are, respectively, given as

$$\rho_{\min} = \frac{e^{-\sigma_X} - 1}{\sqrt{(e-1)(e^{\sigma_X^2} - 1)}}, \;\; \rho_{\max} = \frac{e^{\sigma_X} - 1}{\sqrt{(e-1)(e^{\sigma_X^2} - 1)}}$$

We note that $\rho_{\min} < 0$ and $\rho_{\max} > 0$ for all $\sigma_X > 0$. As σ_X increases, $\rho_{\max}$ first increases (until $\sigma_X = 1$ which results in the highest attainable correlation, i.e., $\rho_{\max} = 1$) and then decreases to zero while $\rho_{\min}$

monotonically increases to zero [14, Fig. 5.8]. The attainable correlation values in the interval $[\rho_{\min}, \rho_{\max}]$ are mostly positive for all values of $\sigma_X > 0$. Due to the asymmetry between $\rho_{\min}$ and $\rho_{\max}$, $\overline{\rho}$ is not necessarily zero and the results developed in Case (a) of Proposition 1 can be employed in applications involving linear estimation of lognormal random variables under correlation uncertainty.

An example employing lognormal random variables is practically relevant in wireless communications since log-normal distribution is used to model shadowing effects [17]. More specifically, it may be desirable to estimate the shadowing at a particular location based on the knowledge of the shadowing at another nearby location. It is well-known that the autocovariance function of the shadowing values between two locations separated by a distance δ is given by $A(\delta) = \sigma^2_{\phi_{\rm dB}} e^{-\delta/d_c}$, where d_c is the decorrelation distance at which autocovariance decreases by a factor of e and $\sigma^2_{\phi_{\rm dB}}$ denotes the variance of lognormal shadowing [17]. Although this formula can be employed to provide an estimate for the correlation coefficient of shadowing between two locations, uncertainty can be introduced due to imprecise knowledge of d_c. Hence, the proposed framework becomes relevant in this practical scenario. The mean correlation coefficient, $\overline{\rho}$, can be computed from $A(\delta)$ along with the bounds of the attainable correlation values, $\rho_{\min}$ and $\rho_{\max}$, as explained above.

We start by investigating the effect of the constraint parameter. The lognormal setting explained in the previous paragraph is considered first. A uniform PDF is assumed for the correlation coefficient, which yields $\overline{\rho} = (\rho_{\min} + \rho_{\max})/2$. It can be shown that $\overline{\rho} > 0$ for all $\sigma_X > 0$. $r \geq \sigma_X^2$ is required for feasibility and, as mentioned above, Case (a) of Proposition 1 applies. Since $\rho^{(1)}_{\max} \leq 0$ and $\overline{\rho} > 0$, we either have $\alpha^{\rm opt} = \overline{\rho}\sigma_X/\sigma_Y$ if $\overline{\rho} \leq \rho^{(2)}_{\min}$ or $\alpha^{\rm opt} = \rho^{(2)}_{\min}\sigma_X/\sigma_Y$ if $\overline{\rho} > \rho^{(2)}_{\min}$. Three different choices are employed for the standard deviation of X, namely, $\sigma_X \in \{1.5, 2, 2.5\}$, and $\sigma_Y = 1$ is assumed throughout this section. The results are presented in Fig. 1. As expected, the average MSE evaluated at the objective function decreases as the constraint on the worst-case MSE is relaxed. Since $\rho^{(2)}_{\min}$ is an increasing function of the constraint threshold r, the condition $\overline{\rho} \leq \rho^{(2)}_{\min}$ will be eventually satisfied as r is kept increasing and the objective function will remain constant at $\sigma_X^2(1 - \overline{\rho}^2)$, the average MSE value corresponding to the solution of the unconstrained optimization problem. In Fig. 1, the performance of the unconstrained optimal linear estimator with respect to the average value of the correlation coefficient is also plotted for comparison since it is the solution to the unconstrained optimization problem.

Next, we consider a case where prior information suggests the presence of negative correlation between the random variables as specified by the parameter values $\rho_{\min} = -0.8$, $\overline{\rho} = -0.6$ and $\rho_{\max} = -0.3$. In

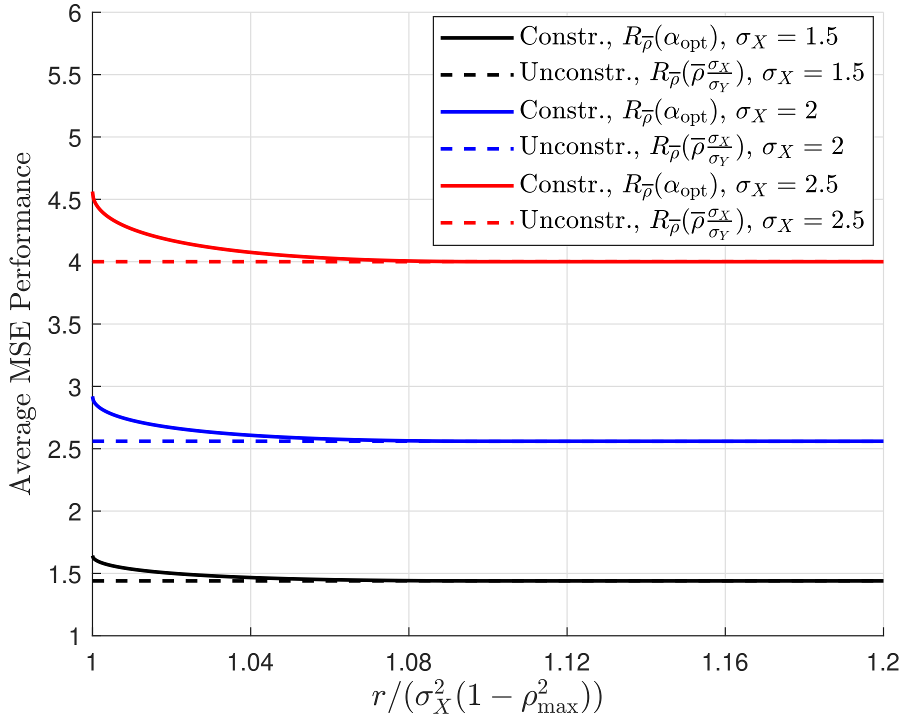

Fig. 2. The value of the objective function evaluated at the solution of the constrained optimization problem (P1) for the negative correlation scenario as a function of the constraint parameter r and different values of σ_X.

this case, $r \geq \sigma_X^2(1-\rho_{\max}^2)$ is required for feasibility. Based on the provided parameter values, it can be seen that $\alpha^{\rm opt} = \rho_{\max}^{(1)}\sigma_X/\sigma_Y$ for $\sigma_X^2(1-\rho_{\max}^2) \leq r < \sigma_X^2$ and $\alpha^{\rm opt} = \overline{\rho}\sigma_X/\sigma_Y$ for $r \geq \sigma_X^2$. The results are presented in Fig. 2.

We also consider the effect of varying σ_X^2, which in turn changes the values of the maximal and the minimal correlations together with the average value under the uniform distribution assumption. It is seen from Fig. 3 that the worst-case MSE can be confined within one-thousandth of σ_X^2 in the expense of a small increase on the average MSE performance.

We conclude this section with an example that demonstrates how the proposed framework ensures that the worst-case MSE is controlled to remain below a feasible threshold value while the optimal linear MMSE estimator based on the mean value of the correlation coefficient fails in this respect. To this end, we assume $\rho_{\min} = -0.6$, $\rho_{\max} = 0.9$, $\overline{\rho} = 0.4$ with $r = 1.2\sigma_X^2$, where $\sigma_X^2 = 1$. Case (a) of Proposition 1 applies in this case with $\rho_{\max}^{(1)} \approx -0.105$ and $\rho_{\min}^{(2)} \approx 0.148$. Since $\overline{\rho} > \rho_{\min}^{(2)}$, the optimal linear estimation coefficient is obtained as $\alpha^{\rm opt} = \rho_{\min}^{(2)}\sigma_X/\sigma_Y$. In Fig. 4, the MSE performance of the optimal estimator, i.e., $R_\rho(\alpha^{\rm opt})$, is plotted as a function of the true correlation coefficient ρ together with the MSE performance of the optimal estimator for the unconstrained problem, i.e., $R_\rho(\overline{\rho}\sigma_X/\sigma_Y)$. As seen from the figure, the constraint on the MSE is violated by the estimator employing the mean

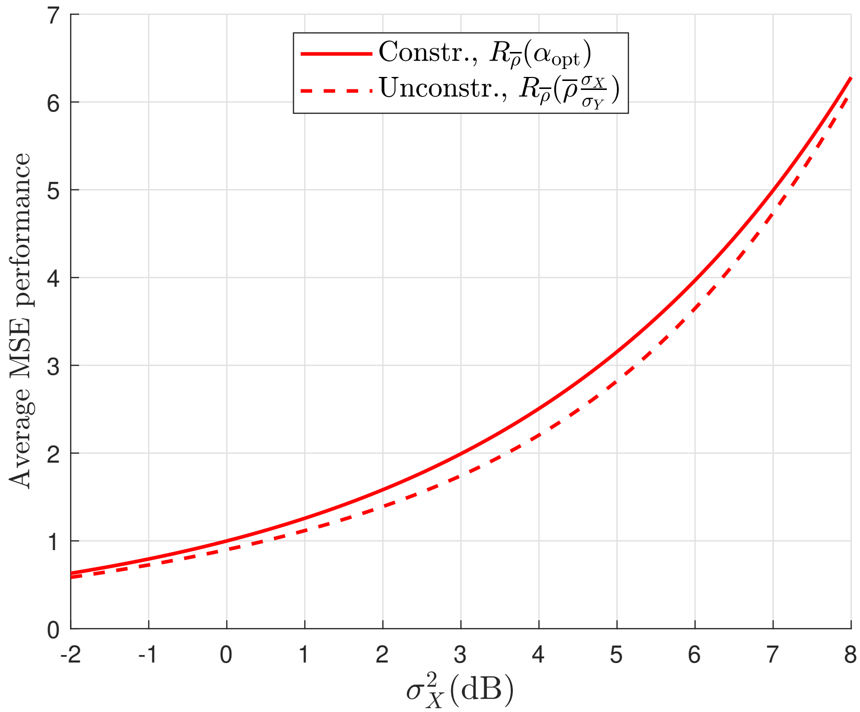

Fig. 3. The values of the objective function evaluated at the solutions of the unconstrained and the constrained optimization problems for the lognormal setting as a function of σ_X. The constraint threshold is fixed at $1.001\sigma_X^2$.

value of the correlation coefficient (i.e., the optimal estimator for the unconstrained case) for all values of $\rho < 0.05$ while the proposed estimator never exceeds the MSE constraint. On the other hand, the average MSE performance (i.e., the MSE performance at the mean value of the correlation coefficient, which is $\overline{\rho} = 0.4$) of the estimator corresponding to the unconstrained case is superior to that of the proposed estimator corresponding to the constrained problem. Hence, a robustness in MSE over the range of possible correlation coefficients is achieved at the expense of higher MSE performance at the expected value of the correlation coefficient. Nevertheless, among all linear estimators that satisfy the worst-case MSE constraint, the proposed solution is the one that achieves the minimum estimation MSE at the mean value of the correlation coefficient.

It is also of interest to compare the performance of the proposed solution with the trivial estimator $\hat{X} = \mu_X$ corresponding to $\alpha = 0$ (hence, observation Y is not used), which has constant MSE equal to σ_X^2 over the range of all possible correlation coefficients. For the considered example, $r > \sigma_X^2$, hence $\hat{X} = \mu_X$ is a feasible estimator. As expected, the MSE of the proposed solution at the mean value $\rho = 0.4$ is lower than that of the trivial estimator. Furthermore, since the MSE performance of the proposed estimator varies linearly with the value of the true correlation coefficient ρ, its MSE performance beats that of the trivial estimator for all $\rho > 0.0875$.

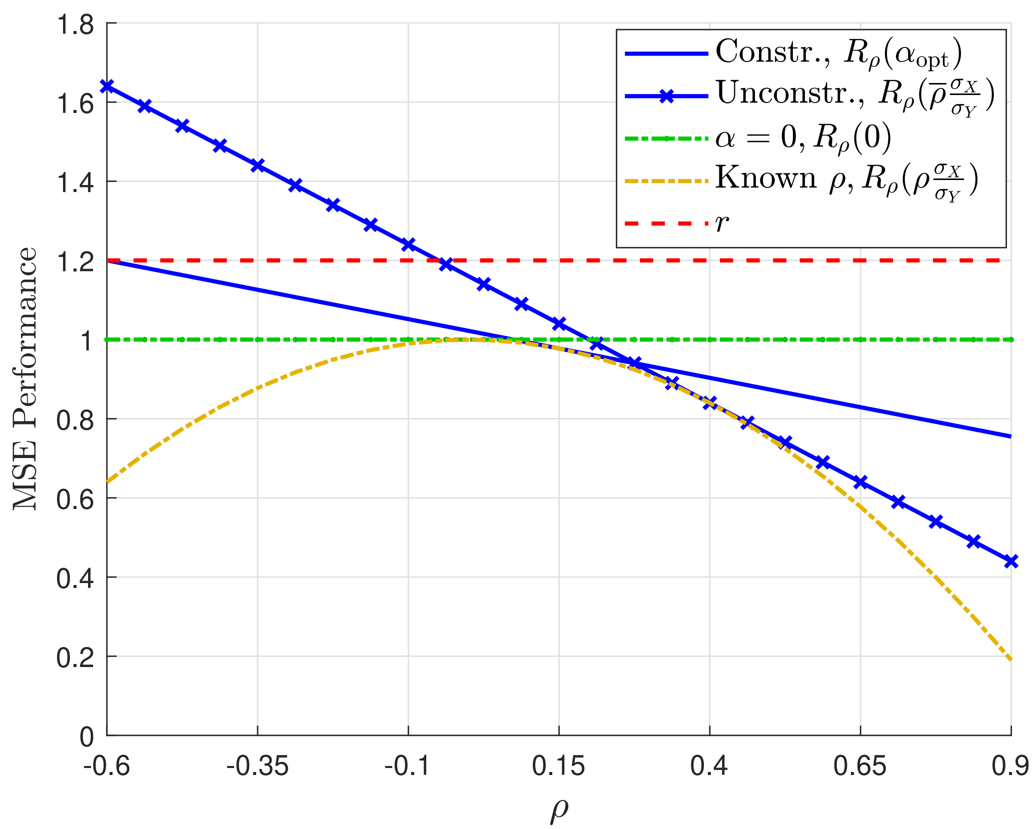

Fig. 4. MSE performance of the optimal estimators for the constrained and unconstrained optimization problems as a function of the true value of the correlation coefficient ρ.

Lastly, a simple, yet tight, lower bound is provided by considering optimal LMMSE estimation based on the true value of the correlation coefficient, i.e., by assuming that ρ is perfectly known, which yields $R_\rho(\rho\frac{\sigma_X}{\sigma_Y}) = \sigma_X^2(1-\rho^2), \forall\rho \in [\rho_{\min}, \rho_{\max}]$. The MSE performance of the proposed optimal estimator, i.e., $R_\rho(\alpha_{\text{opt}})$, is tangent to the lower bound at $\rho = \rho_{\min}^{(2)} \approx 0.148$ which lies in between 0 and $\overline{\rho} = 0.4$, exemplifying the operation of the proposed estimator as a shrinkage operator.

V. Extension to Vector-Valued Observation

In this section, we present a characterization of the optimal solution when the observation is vector-valued, i.e., $\boldsymbol{Y} = (Y_1, \dots, Y_n) \in \mathbb{R}^n$ for some integer $n > 1$.[1] In this case, a linear unbiased estimator can be written in the following form:

$$\hat{X}_{\boldsymbol{\alpha}}(\boldsymbol{Y}) = \mu_X + \boldsymbol{\alpha}^T(\boldsymbol{Y} - \boldsymbol{\mu}_Y), \tag{18}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^n$ is the coefficient vector subject to design. With a change of variables, we substitute $\boldsymbol{\alpha} = \sigma_X(\mathbf{D}_Y)^{-1}\boldsymbol{\varrho}$, where $\mathbf{D}_Y$ is the $n \times n$ diagonal matrix with its i-th diagonal entry equal to the standard

[1]All vectors are column vectors. Row vectors are denoted by taking the transpose, i.e., $(\cdot)^T$.

deviation of the i-th observation Y_i, i.e., $[\mathbf{D}_Y]_{i,i} = \sigma_{Y_i}$, and $\boldsymbol{\varrho} = (\varrho_1, \ldots, \varrho_n)$ is the new parameter vector. Similar to (8), it can be shown that the corresponding MSE is expressed as

$$R_{\boldsymbol{\rho}}(\boldsymbol{\varrho}) = \sigma_X^2 \left(1 - 2\boldsymbol{\rho}^T \boldsymbol{\varrho} + \boldsymbol{\varrho}^T \mathbf{R}_Y \boldsymbol{\varrho}\right) , \tag{19}$$

where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)$ is vector of true cross-correlation coefficients between X and $\boldsymbol{Y}$ with $\rho_i = \mathrm{Cov}(X, Y_i)/(\sigma_X \sigma_{Y_i})$, and $\mathbf{R}_Y$ is the $n \times n$ correlation coefficient matrix of $\boldsymbol{Y}$ with $[\mathbf{R}_Y]_{i,j} = \mathrm{Cov}(Y_i, Y_j)/(\sigma_{Y_i} \sigma_{Y_j})$. We assume that the variance of each Y_i is finite and non-zero, and the covariance matrix of $\boldsymbol{Y}$ is positive definite, hence, so is $\mathbf{R}_Y$. The optimal linear MMSE estimator is given as $\boldsymbol{\alpha}^* = \sigma_X (\mathbf{D}_Y)^{-1} \boldsymbol{\varrho}^*$, where $\boldsymbol{\varrho}^* = (\mathbf{R}_Y)^{-1} \boldsymbol{\rho}$ is the minimizer of $R_{\boldsymbol{\rho}}(\boldsymbol{\varrho})$. We consider the case where $\boldsymbol{\rho}$ is unknown. In order to model the uncertainty in the cross-correlation coefficient vector, it is assumed that $\boldsymbol{\rho}$ is distributed with PDF $f(\boldsymbol{\rho})$ over a convex and compact set $\Theta \subset [-1, 1]^{\otimes n}$, where $\otimes n$ denotes the n-th Cartesian product. The mean value is denoted as $\mathbb{E}[\boldsymbol{\rho}] = \overline{\boldsymbol{\rho}}$. The proposed restricted Bayesian linear estimation problem for vector-valued observation case can be expressed as

$$\begin{aligned} \min_{\boldsymbol{\varrho} \in \mathbb{R}^n} \quad & \mathbb{E}\left[R_{\boldsymbol{\rho}}(\boldsymbol{\varrho})\right] \\ \text{subject to} \quad & R_{\boldsymbol{\rho}}(\boldsymbol{\varrho}) \leq r \qquad \forall \boldsymbol{\rho} \in \Theta \, . \end{aligned} \tag{P3}$$

By the linearity of $R_{\boldsymbol{\rho}}(\boldsymbol{\varrho})$ in $\boldsymbol{\rho}$, we obtain $\mathbb{E}[R_{\boldsymbol{\rho}}(\boldsymbol{\varrho})] = R_{\overline{\boldsymbol{\rho}}}(\boldsymbol{\varrho})$. Expanding the terms in (P3), we equivalently get

$$\begin{aligned} \min_{\boldsymbol{\varrho} \in \mathbb{R}^n} \quad & \boldsymbol{\varrho}^T \mathbf{R}_Y \boldsymbol{\varrho} - 2\overline{\boldsymbol{\rho}}^T \boldsymbol{\varrho} + 1 \\ \text{subject to} \quad & \max_{\boldsymbol{\rho} \in \Theta} \; \boldsymbol{\varrho}^T \mathbf{R}_Y \boldsymbol{\varrho} - 2\boldsymbol{\rho}^T \boldsymbol{\varrho} + 1 \leq r/\sigma_X^2 \, . \end{aligned} \tag{P4}$$

The maximization performed in the constraint function of (P4) has an interesting interpretation in convex analysis. Since $R_{\boldsymbol{\rho}}(\boldsymbol{\varrho})$ is linear in $\boldsymbol{\rho}$, the optimization performed in the constraint function can be expressed as

$$\max_{\boldsymbol{\rho} \in \Theta} \; R_{\boldsymbol{\rho}}(\boldsymbol{\varrho}) = \min_{\boldsymbol{\rho} \in \Theta} \; \boldsymbol{\varrho}^T \boldsymbol{\rho} \, . \tag{20}$$

The solution to (20) is obtained by minimizing a linear function over a convex set. By the supporting hyperplane theorem [18, Section 2.5.2], it yields a point on the boundary of the convex set, denoted with $\boldsymbol{\rho}_{\mathrm{LF}}$. The subscript indicates that the resulting correlation coefficient vector is the least favorable one as it maximizes the estimation MSE for the employed linear estimation coefficient $\boldsymbol{\varrho}$, which corresponds to the normal vector of the hyperplane that is tangent to the convex set Θ at the point $\boldsymbol{\rho}_{\mathrm{LF}}$. In other words, all cross-correlation coefficient vectors in the set Θ satisfy

$$\boldsymbol{\varrho}^T(\boldsymbol{\rho} - \boldsymbol{\rho}_{\mathrm{LF}}) \geq 0 \quad \forall\, \boldsymbol{\rho} \in \Theta\,, \tag{21}$$

and $\boldsymbol{\rho}_{\mathrm{LF}}$ can be computed either analytically or using standard tools from convex optimization. The dependence of $\boldsymbol{\rho}_{\mathrm{LF}}$ on the normal vector $\boldsymbol{\varrho}$ can be made explicit by denoting it as $\boldsymbol{\rho}_{\mathrm{LF}}(\boldsymbol{\varrho})$.

We note that the objective in (P4) is a convex function of $\boldsymbol{\varrho}$ since $\mathbf{R}_Y > 0$. Likewise, the convexity of the constraint function follows from the fact that the pointwise maximum of a set of convex functions is convex [18, Section 3.2.3]. Hence, the optimization problem in (P4) is convex. The following proposition characterizes the optimal solution for the case of vector observations.

***Proposition 2**: The optimal solution to* (P3)*, and equivalently to* (P4)*, is characterized as follows:*

Case 1. *Let* $\boldsymbol{\rho}^*_{\mathrm{LF}} = \arg\min_{\boldsymbol{\rho}\in\Theta}\ \overline{\boldsymbol{\rho}}^T(\mathbf{R}_Y)^{-1}\boldsymbol{\rho}$*. If* $R_{\boldsymbol{\rho}^*_{\mathrm{LF}}}((\mathbf{R}_Y)^{-1}\overline{\boldsymbol{\rho}}) \leq r$*, then* $\boldsymbol{\varrho}^{\mathrm{opt}} = (\mathbf{R}_Y)^{-1}\overline{\boldsymbol{\rho}}$*. Otherwise,*

Case 2. *The optimal solution* $\boldsymbol{\varrho}^{\mathrm{opt}}$ *satisfies:*

i. $\boldsymbol{\varrho}^{\mathrm{opt}} = (\mathbf{R}_Y)^{-1}\left(\lambda\boldsymbol{\rho}^*_{\mathrm{LF}} + (1-\lambda)\overline{\boldsymbol{\rho}}\right)$ *for* $\lambda \in (0, 1]$*, where* $\boldsymbol{\rho}^*_{\mathrm{LF}} = \arg\min_{\boldsymbol{\rho}\in\Theta}\ (\boldsymbol{\varrho}^{\mathrm{opt}})^T\boldsymbol{\rho}$*.*

ii. $R_{\boldsymbol{\rho}^*_{\mathrm{LF}}}(\boldsymbol{\varrho}^{\mathrm{opt}}) = r$*. More explicitly,* $(\boldsymbol{\varrho}^{\mathrm{opt}})^T\mathbf{R}_Y\boldsymbol{\varrho}^{\mathrm{opt}} - 2(\boldsymbol{\rho}^*_{\mathrm{LF}})^T\boldsymbol{\varrho}^{\mathrm{opt}} = r/\sigma_X^2 - 1$*.*

Then, the optimal coefficient vector is formed as $\boldsymbol{\alpha}^{\mathrm{opt}} = \sigma_X(\mathbf{D}_Y)^{-1}\boldsymbol{\varrho}^{\mathrm{opt}}$*.*

Proof: The proof is based on the observations mentioned in the previous paragraphs and uses the fact that KKT conditions are necessary and sufficient. Case 1 checks whether the solution to the unconstrained optimization problem satisfies the constraint on the worst-case MSE. If not, we proceed with Case 2. In this case, forming the Lagrangian and setting the derivative equal to zero yields Case 2-i. It basically states that the optimal solution is a convex combination of the mean cross-correlation vector $\overline{\boldsymbol{\rho}}$ and the least favorable cross-correlation vector $\boldsymbol{\rho}^*_{\mathrm{LF}}$. As described above, the least favorable cross-correlation vector $\boldsymbol{\rho}^*_{\mathrm{LF}}$ is a point on the boundary of the set Θ and there exists a supporting hyperplane passing through this point defined by $\{\boldsymbol{\rho} \in \mathbb{R}^n : (\boldsymbol{\varrho}^{\mathrm{opt}})^T(\boldsymbol{\rho} - \boldsymbol{\rho}^*_{\mathrm{LF}}) = 0\}$. Hence, it holds that $(\boldsymbol{\varrho}^{\mathrm{opt}})^T(\boldsymbol{\rho} - \boldsymbol{\rho}^*_{\mathrm{LF}}) \geq 0$ for all $\boldsymbol{\rho} \in \Theta$. Lastly in Case 2-ii, from complementary slackness and positive Lagrange multiplier, it is concluded that the constraint must be satisfied with equality, i.e., $R_{\boldsymbol{\rho}^*_{\mathrm{LF}}}(\boldsymbol{\varrho}^{\mathrm{opt}}) = r$. □

VI. CONCLUSION

In this paper, we proposed a restricted Bayes approach for linear estimation of a scalar random parameter based on a scalar observation when there exists correlation uncertainty between the parameter and the observation. We derived the optimal linear estimator that minimizes the average MSE under a constraint on the worst-case MSE by utilizing the statistical knowledge on the correlation coefficient. We obtained a closed-form expression for the proposed optimal linear estimator and evaluated its performance via numerical examples which illustrated its benefits in various scenarios. We also extended the results to the case of vector-valued observation.

As future work, the proposed restricted Bayesian approach for linear estimation under correlation uncertainty can be applied to the decentralized estimation problem where the aim is to optimally combine correlated local state estimates at the fusion center under uncertainty about their cross-correlation. While robust fusion algorithms that are based on the uncertainty sets for the error covariance matrix of local state estimates exist in the relevant distributed estimation literature, a joint estimation approach that minimizes the average estimation error with respect to a nomimal distribution while controlling the worst-case fused MSE is not available to the best of our knowledge. Another possible venue for future work is the design of theoretical bounds for average estimation error under uncertainty on the joint distribution (e.g., unknown correlation coefficient in the case of linear estimators).

APPENDIX
PROOF OF PROPOSITION 1

We analyze the solution under the three cases stated in the lemma. First, we specify the range of feasible values of $\varrho \in \mathbb{R}$ in all the cases. In Case 1 (i.e., $\rho_{\min} < 0 < \rho_{\max}$), it is required to have $r \geq \sigma_X^2$ for feasibility. Under this condition, the range of feasible values of ϱ can be specified explicitly. For $\varrho \geq 0$, the first inequality constraint in (P2) is satisfied for $\varrho \in [0, \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}]$ and for $\varrho \leq 0$, the second inequality constraint in (P2) is satisfied for $\varrho \in [\rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}, 0]$. Hence, the range of feasible values of ϱ under Case 1 is specified as

$$\text{Case 1: } \varrho \in \left[\rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}, \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}\right]. \tag{22}$$

In Case 2 (i.e., $0 \leq \rho_{\min} < \rho_{\max}$), it is required to have $r \geq \sigma_X^2(1 - \rho_{\min}^2)$ for feasibility. This case is divided into two subcases: (2a) $\sigma_X^2(1 - \rho_{\min}^2) \leq r \leq \sigma_X^2$, and (2b) $r \geq \sigma_X^2$. In case (2a), the second inequality constraint in (P2) does not produce any feasible $\varrho < 0$, and we obtain the range of feasible values of ϱ based solely on the first inequality constraint as

$$\text{Case 2a: } \varrho \in \left[\rho_{\min} - \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}, \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}\right]. \tag{23}$$

In Case (2b), both inequality constraints in (P2) contribute to the set of feasible values of ϱ, which yields the same result as given in Case 1. In Case 3 (i.e., $\rho_{\min} < \rho_{\max} \leq 0$), it is required to have $r \geq \sigma_X^2(1-\rho_{\max}^2)$

for feasibility. Again, this case can be decomposed into two subcases: (3a) $\sigma_X^2(1-\rho_{\max}^2) \le r \le \sigma_X^2$, and (3b) $r \ge \sigma_X^2$. By a similar analysis, the range of feasible values of ϱ under Case 3a is obtained as

$$\text{Case 3a: } \varrho \in \left[\rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}, \rho_{\max} + \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}\right], \tag{24}$$

and Case (3b) produces the same range of feasible values as given in Case 1. Once the set of all feasible values of ϱ is determined under all three cases, the optimal value of ρ, denoted with $\rho^{\rm opt}$, is obtained as the value that minimizes the objective function given in (P2). Since the objective is a convex quadratic function of ϱ with a unique minimum at $\varrho = \overline{\rho}$, the optimal value $\rho^{\rm opt}$ is given by the value of ρ in the feasible set that is closest to $\overline{\rho}$. Under Cases 1, (2b), and (3b), which are all characterized by the condition $r \ge \sigma_X^2$, if $\overline{\rho}$ in an element of the feasible set given in (22), then $\rho^{\rm opt} = \overline{\rho}$. If $\overline{\rho}$ does not belong to the feasible set, $\rho^{\rm opt} = \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}$ if $\overline{\rho} > \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}$; and $\rho^{\rm opt} = \rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}$ if $\overline{\rho} < \rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}$. Under Case (2a), which is characterized by the condition $\sigma_X^2(1-\rho_{\min}^2) \le r \le \sigma_X^2$, if $\overline{\rho}$ in an element of the feasible set given in (23), then $\rho^{\rm opt} = \overline{\rho}$. If $\overline{\rho}$ does not belong to the feasible set, $\rho^{\rm opt} = \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}$ when $\overline{\rho} > \rho_{\min} + \sqrt{\rho_{\min}^2 + r/\sigma_X^2 - 1}$. That covers all possibilities for Case (2a) since $\rho_{\min} < \overline{\rho} < \rho_{\max}$ by definition. Under Case (3a), which is characterized by the condition $\sigma_X^2(1-\rho_{\max}^2) \le r \le \sigma_X^2$, if $\overline{\rho}$ in an element of the feasible set given in (24), then $\rho^{\rm opt} = \overline{\rho}$. If $\overline{\rho}$ does not belong to the feasible set, $\rho^{\rm opt} = \rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}$ when $\overline{\rho} < \rho_{\max} - \sqrt{\rho_{\max}^2 + r/\sigma_X^2 - 1}$. Again, all possibilities are covered for Case (3a). Substituting $\alpha^{\rm opt} = \rho^{\rm opt}\sigma_X/\sigma_Y$, using the fact that $\rho_{\min} < \overline{\rho} < \rho_{\max}$, and arranging the results according to these cases, the results presented in the proposition are obtained. □

REFERENCES

[1] H. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer-Verlag, New York, 1994.

[2] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused-track covariance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 6, pp. 803–805, Nov. 1986.

[3] K. Chang, R. Saha, and Y. Bar-Shalom, "On optimal track-to-track fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 4, pp. 1271–1276, Oct. 1997.

[4] R. Forsling, A. Hansson, F. Gustafsson, Z. Sjanic, J. Löfberg, and G. Hendeby, "Conservative linear unbiased estimation under partially known covariances," *IEEE Trans. Sig. Process.*, vol. 70, pp. 3123–3135, June 2022.

[5] M. A. Bakr and S. Lee, "Distributed multisensor data fusion under unknown correlation and data inconsistency," *Sensors*, vol. 17, no. 11, Nov. 2017, Art. No. 2472.

[6] 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, Jan. 2015.

[7] M. Reinhardt, B. Noack, M. Baum, and U. D. Hanebeck, “Analysis of set-theoretic and stochastic models for fusion under unknown correlations,” in *Proc. 14th Int. Conf. Inf. Fusion*, Chicago, IL, USA, July 5–8 2011.

[8] L. Kaplan, W. Blair, and Y. Bar-Shalom, “Simulations studies of multisensor track association and fusion methods,” in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, March 4–11 2006.

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

[10] X. Qu, J. Zhou, E. Song, and Y. Zhu, “Minimax robust optimal estimation fusion in distributed multisensor systems with uncertainties,” *IEEE Signal Process. Lett.*, vol. 17, no. 9, pp. 811–814, May 2010.

[11] Y. Gao, X. R. Li, and E. Song, “Robust linear estimation fusion with allowable unknown cross-covariance,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 9, pp. 1314–1325, Oct. 2016.

[12] J. Tang, J. Zhou, and Y. Rong, “Estimation fusion for distributed multi-sensor systems with uncertain cross-correlations,” *Int. J. Syst. Sci.*, vol. 50, no. 7, pp. 1378–1387, May 2019.

[13] J. L. H. Jr. and E. L. Lehmann, “The use of previous experience in reaching statistical decisions,” *Ann. Math. Stat.*, vol. 23, no. 3, pp. 396 – 407, Sept. 1952.

[14] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2015.

[15] M. Fréchet, “Sur les tableaux de corrélation dont les marges sont données,” *Ann. Univ. Lyon, Sect. A: Sci. Math. Astr.*, vol. 9, pp. 53–77, 1951.

[16] B. Efron and T. Hastie, *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, ser. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2016.

[17] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.