Bayesian Minimax Estimation of the Normal Model With Incomplete Prior Covariance Matrix Specification

Duc-Son Pham, *Member, IEEE*, Hung H. Bui, and Svetha Venkatesh, *Senior Member, IEEE*

***Abstract*—This work addresses the issue of Bayesian robustness in the multivariate normal model when the prior covariance matrix is not completely specified, but rather is described in terms of positive semi-definite bounds. This occurs in situations where, for example, the only prior information available is the bound on the diagonal of the covariance matrix derived from some physical constraints, and that the covariance matrix is positive semi-definite, but otherwise arbitrary. Under the conditional Gamma-minimax principle, previous work by DasGupta and Studden shows that an analytically exact solution is readily available for a special case where the bound difference is a scaled identity. The goal in this work is to consider this problem for general positive definite matrices. The contribution in this paper is a theoretical study of the geometry of the minimax problem. Extension of previous results to a more general case is shown and a practical algorithm that relies on semi-definite programming and the convexity of the minimax formulation is derived. Although the algorithm is numerically exact for up to the bivariate case, its exactness for other cases remains open. Numerical studies demonstrate the accuracy of the proposed algorithm and the robustness of the minimax solution relative to standard and recently proposed methods.**

***Index Terms*—Bayesian point estimate, gamma-minimax, minimax estimator, normal model, prior uncertainty, robust Bayesian analysis, shrinkage method.**

I. Introduction

We consider the estimation problem of random parameters $\boldsymbol{\theta}$ in the multivariate normal model under the Bayesian paradigm

$$\mathbf{y}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{C}\boldsymbol{\theta}, \mathbf{R_{nn}}). \tag{1}$$

Here, both the measurement noise covariance matrix $\mathbf{R}_{\boldsymbol{nn}}$ and the design matrix $\mathbf{C} \in \mathbb{R}^{N\times K}$ are known, and the random parameter $\boldsymbol{\theta} \in \mathbb{R}^K$ $(K < N)$ follows a conjugate prior given as:

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu_\theta}, \mathbf{R}_{\boldsymbol{\theta\theta}}). \tag{2}$$

Manuscript received July 31, 2008; revised March 02, 2010. Date of current version November 19, 2010.

D.-S. Pham and S. Venkatesh are with the Institute for Multisensor Processing and Content Analysis (IMPCA), Curtin University of Technology, Perth, Western Australia (e-mail: dspham@ieee.org; svetha@cs.curtin.edu.au).

H. H. Bui is with the Artificial Intelligence Center, SRI International, Menlo Park, CA 94025-3493 USA (e-mail: bui@ai.sri.com).

Communicated by F. Hlawatsch, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2080612

When this prior is completely specified and both $\mathbf{R}_{\boldsymbol{\theta\theta}}$ and $\mathbf{R_{nn}}$ are invertible, it is widely known [21] that there exists an analytical Bayesian solution to this model given by

$$\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{y}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}|\mathbf{y}}\right)$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}|\mathbf{y}} = \left(\mathbf{R}_{\boldsymbol{\theta\theta}}^{-1} + \mathbf{C}^T\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{C}\right)^{-1} \tag{3}$$

$$\boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{y}} = \boldsymbol{\mu_\theta} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}|\mathbf{y}}\boldsymbol{\eta},\ \boldsymbol{\eta} = \mathbf{C}^T\mathbf{R}_{\mathbf{nn}}^{-1}(\mathbf{y} - \mathbf{C}\boldsymbol{\mu_\theta}). \tag{4}$$

The posterior mean $\boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{y}}$ minimizes the Bayes risk under the quadratic loss. The analytical solution has led to many widely applicable algorithms, for example, the Kalman filter.

However, complete prior specification may not always be available to a statistician for various reasons, such as, an insufficient number of observations to ensure good domain knowledge, missing and noisy data, or many parameters in large-scale problems [20]. In particular, we consider the situation addressed by Leamer [14] and DasGupta and Studden [9] that the prior belongs to a class of distributions Γ with known mean, and the covariance matrix can only be specified by the upper and lower bounds as

$$\Gamma = \{\pi(\boldsymbol{\theta}) : \pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu_\theta}, \mathbf{R}_{\boldsymbol{\theta\theta}}),\ \ \mathbf{R}_{\boldsymbol{\theta\theta}} \in \mathcal{R}\} \tag{5}$$

$$\mathcal{R} = \{\mathbf{R}_{\boldsymbol{\theta\theta}} : \mathbf{0} \prec \mathbf{R}_{\min} \preceq \mathbf{R}_{\boldsymbol{\theta\theta}} \preceq \mathbf{R}_{\max}\} \tag{6}$$

where $\preceq$ denotes the *Löwner* partial order for symmetric matrices[1] (i.e., $\mathbf{A} \preceq \mathbf{B}$ means $\mathbf{B} - \mathbf{A}$ is positive semi-definite (psd). Similarly, $\mathbf{A} \prec \mathbf{B}$ means $\mathbf{B} - \mathbf{A}$ is strictly positive definite.)

We note that the specification in (6) is not the only unique way to describe the incomplete prior. One may also bound the covariance matrix in terms of a Frobenius norm, for example, as in [12]. Other constraints are also discussed by Eldar and Merhav in [10]. However, we note that the constraint (6) implies constraints on the trace, determinant, and norm but not the converse. Accordingly, if the constraints in [10] are imposed, an analytical minimax solution can be derived from the subsequent results of this work (see Appendix A). The incomplete specification is only for the covariance matrix due to the assumption that, in practice, eliciting the prior mean rather than higher moments is easier for a statistician. More arguments for the constraint (6) are discussed in [9] and references therein.

[1]For a review of generalized inequalities which the *Löwner* partial order is a special case, see [5, p. 43].

The above incomplete prior specification problem can be tackled by different approaches under the Bayesian paradigm (a thorough discussion is given in Berger's textbook [3, Ch. 4]). For example, parametric empirical Bayes treats the variances as hyperparameters that can be estimated from the data. However, there is a problem with passing the uncertainty in the hyperparameter estimates to the final decision for small and moderate size problems unless one develops approximations (such as [17]) that account for that uncertainty [3, p. 170]. A full Bayesian treatment (i.e., hierarchical Bayes analysis) places hyperpriors on top of the prior distribution and performs nested integrations to obtain the final estimates. Hierarchical Bayes is often robust [3], but the main drawback is the computational burden, unless one uses approximations by assuming other parametric forms for the hyperpriors. We note that the estimation problem of normal means under both empirical and hierarchical Bayes analysis is well documented in [3, Ch. 4].

In this work, we are interested in an approach within the Bayesian paradigm that is concerned more with the sensitivity of the analysis to possible mis-specification of the prior distribution. Given a class Γ of plausible prior distributions that a statistician can elicit, robust Bayesian analysis finds a decision rule that is optimized for the *worst-case* scenarios. In other words, robust Bayesian analysis ensures that the decision made is guarded against the least favorable prior in Γ[3], [24]. Such a minimax approach is used by many statistical decision procedures [2], [6], [8]. Even in the robust Bayesian frameworks, there are numerous ways to design procedures that possess either prior or posterior robustness (see [3, Ch. 4.7] for more details). For the problem under consideration, we propose to follow the conditional Γ-minimax framework that was first posed by Watson [25] and subsequently studied in [4], [9], and [24]. The key difference with respect to the classical Γ-minimax framework [3], [24] is that it does not integrate over unobserved data to obtain robustness. We also note that there are other variations of the Γ-minimax principle such as the Γ-minimax regret [3], [24]. Our aim in this work is not to advocate any particular Bayesian approach[2]. Rather, our main goal is to extend the previous work by DasGupta and Studden [9], who have solved this problem for a special case. More specifically, we place no restriction on the bounds of the covariance matrix. Our first major contribution is a theoretical result (Theorem 3) that indicates when an analytical minimax solution is available and it covers previous results by DasGupta and Studden [9] as a special case. The second contribution is an algorithm to compute the minimax solution when such an analytical result is not available. This algorithm exploits the geometry of the minimax problem and is based on semi-definite programming. Our theoretical analysis of the algorithm indicates that the algorithm is numerically exact for the bivariate case (i.e., $K = 2$). Further numerical studies suggest that it is also numerically exact for $K = 3$. However, its exactness for $K \geq 3$ is an open question.

The paper is organized as follows. In Section II, we detail the Γ-minimax approach and the geometry of the minimax problem. We then present a theoretical study of this problem and specify the conditions for when an analytical solution is available, thereby extending the previous result of DasGupta and Studden [9]. When the analytical solution does not exist, we provide a numerical algorithm that is based on semi-definite programming and exploits the convexity of the minimax formulation, to compute the minimax solution. In Section III, we perform numerical studies to investigate the accuracy and compare the proposed method to some commonly used Bayesian and recently proposed methods [10]. Concluding remarks are given in Section IV.

[2]A complete discussion on robust Bayesian analysis is given in [3] and a comparison between some robust Bayesian methods is given in [9]

Notation: The field of real numbers is denoted as $\mathbb{R}$, whilst $\mathbb{R}^K$ denotes a K-dimensional vector space over $\mathbb{R}$. Vector quantities are in lower-case bold-face (e.g., $\mathbf{x}, \boldsymbol{\mu}$), matrix quantities are in upper-case bold-face (e.g., $\mathbf{R}, \boldsymbol{\Omega}$), set quantities are in calligraphic letters (e.g., $\mathcal{S}, \mathcal{R}$), $\mathsf{E}[\cdot]$ denotes expectation, $\mathsf{tr}[\cdot]$ denotes trace, $\mathsf{bd}(\cdot)$ denotes the boundary of a set. For technical clarity, the statement $\mathbf{A} \precsim \mathbf{B}$ means $\mathbf{B}-\mathbf{A}$ is positive semi-definite *singular* (i.e., at least one eigenvalue of $\mathbf{B} - \mathbf{A}$ is zero). Finally, all technical proofs are detailed in the Appendices.

II. The Γ-Minimax Solution

We now detail the Γ-minimax approach for the problem of estimating $\boldsymbol{\theta}$ given the condition (5).

Following the conditional Γ-minimax principle ([9], [25]), we are interested in a statistical decision rule $\delta(\mathbf{y}) \rightarrow \hat{\boldsymbol{\theta}}$ that minimizes the worst-case conditional risk

$$\hat{\boldsymbol{\theta}}(\mathbf{y}) = \arg \inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} \mathsf{E}_{F^{\pi}(\boldsymbol{\theta}|\mathbf{y})} \left[L(\boldsymbol{\theta}, \delta(\mathbf{y})) \right]. \tag{7}$$

Here, $\mathcal{D}$ denotes the class of nonrandomized decision rules, $F^{\pi}(\boldsymbol{\theta}|\mathbf{y})$ denotes the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{y}$, and the loss function is quadratic (i.e., $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2$). The ultimate goal of this work is to solve the minimax problem (7).

We note in particular that the minimax solution in (7) is conditional on the observation $\mathbf{y}$. This is different to the classical Γ-minimax approach [3] where an integration over the unobserved data is performed. A previous study by Betrò and Ruggeri [4] establishes a general set of conditions under which the conditional Γ-minimax principle is admissible and the solution is unique.

For notational simplicity, we drop the posterior subscript when referring to the posterior covariance matrix (i.e., we write $\boldsymbol{\Sigma}$ instead of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}|\mathbf{y}}$). Furthermore, we introduce $\boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{\theta}|\mathbf{y}} - \boldsymbol{\mu}_{\boldsymbol{\theta}}$ (note that $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is already known) so that from (4) we have $\boldsymbol{\mu} = \boldsymbol{\Sigma}\boldsymbol{\eta}$. Next, we introduce $\boldsymbol{\Omega} = \mathbf{C}^T \mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{C}$ and assume that $\boldsymbol{\Omega} \succ \mathbf{0}$. It follows that the posterior covariance matrix can be written as $\boldsymbol{\Sigma} = (\mathbf{R}_{\boldsymbol{\theta\theta}}^{-1} + \boldsymbol{\Omega})^{-1}$ and it depends only on $\mathbf{R}_{\boldsymbol{\theta\theta}}$. With this new variable, the original minimax problem (7) over $\boldsymbol{\theta}$ and $\mathbf{R}_{\boldsymbol{\theta\theta}}$ is now reparameterized in terms of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ instead. As the prior covariance matrix $\mathbf{R}_{\boldsymbol{\theta\theta}}$ is bounded by $\mathcal{R}$ defined in (6), it follows that $\boldsymbol{\Sigma}$ is bounded by $\mathcal{S}$, defined as $\mathcal{S} = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma}_{\min} \preceq \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_{\max}\}$, where $\boldsymbol{\Sigma}_{\min} = \left(\mathbf{R}_{\min}^{-1} + \boldsymbol{\Omega}\right)^{-1}$ and $\boldsymbol{\Sigma}_{\max} = \left(\mathbf{R}_{\max}^{-1} + \boldsymbol{\Omega}\right)^{-1}$.

The following result expresses the minimax problem (7) in a more convenient form by explicitly evaluating the conditional risk.

Lemma 1: The minimax problem (7) for the constraints (6) can be equivalently written as

$$\inf_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^K} \quad \sup_{\boldsymbol{\Sigma} \in \mathcal{S}} f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \tag{8}$$

where the re-parameterized conditional risk $f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$ is given by

$$f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) = \| \hat{\boldsymbol{\theta}} - \boldsymbol{\mu} \|^2 + \mathrm{tr}[\boldsymbol{\Sigma}]. \tag{9}$$

The proof is detailed in Appendix B. Hereafter, we define the max problem of (8) as

$$\sup_{\boldsymbol{\Sigma} \in \mathcal{S}} f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \tag{10}$$

and the min problem of (8) as

$$\inf_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^K} g(\hat{\boldsymbol{\theta}}) \tag{11}$$

where $g(\hat{\boldsymbol{\theta}}) \triangleq \sup_{\boldsymbol{\Sigma} \in \mathcal{S}} f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$.

A. Analytical Minimax Solution

We first revisit and strengthen the result on the ellipsoidal bound of the conditional mean $\boldsymbol{\mu}$ by Leamer [14], as this bound is important for the derivation of the minimax solution. We then reduce the set of possible candidates for the inner maximization problem with respect to $\boldsymbol{\Sigma}$ in (8). This allows us to characterize the properties of the minimax solution. We show that the minimax solution can be obtained analytically when a certain condition on the principal eigenvalues of the bound difference matrix is met (see Theorem 3). The result generalizes that of DasGupta and Studden [9], which covers a more special case.

1) Variation of the Conditional Mean:

Lemma 2: The conditional mean $\boldsymbol{\mu} \in \mathcal{E}$ if and only if $\boldsymbol{\Sigma} \in \mathcal{S}$, where $\mathcal{E}$ is an ellipsoid given by

$$\mathcal{E} = \{\boldsymbol{\mu} : (\boldsymbol{\mu} - \boldsymbol{\mu}_c)^T \boldsymbol{\Phi} (\boldsymbol{\mu} - \boldsymbol{\mu}_c) \leq c_{\mathcal{E}}\} \tag{12}$$

and

$$\boldsymbol{\mu}_c = \frac{1}{2}(\boldsymbol{\Sigma}_{\max} + \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta} \tag{13}$$

$$\boldsymbol{\Phi} = (\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min})^{-1} \tag{14}$$

$$c_{\mathcal{E}} = \frac{1}{4}\boldsymbol{\eta}^T(\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta}. \tag{15}$$

The proof is detailed in Appendix C.

Further Geometrical Interpretation: We further review some geometrical results on ellipsoidal bounds derived in [14] that will be used subsequently for the derivation of the minimax estimator (see Fig. 1 for an illustration of the 2-D case).

- If only the positive definite condition is imposed, i.e., $\mathbf{R}_{\boldsymbol{\theta\theta}} \succ \mathbf{0}$, then $\boldsymbol{\mu}$ is within an ellipsoid $\mathcal{E}_0$ that goes through the origin and whose center is at $\frac{\boldsymbol{\mu}_0}{2}$ where $\boldsymbol{\mu}_0 = \boldsymbol{\Omega}^{-1}\mathbf{C}^T\mathbf{R}_{\mathbf{nn}}^{-1}(\mathbf{y} - \mathbf{C}\boldsymbol{\mu}_{\boldsymbol{\theta}})$. Mathematically, the ellipsoid $\mathcal{E}_0$ is defined as

$$\left(\boldsymbol{\mu} - \frac{\boldsymbol{\mu}_0}{2}\right)^T \boldsymbol{\Omega} \left(\boldsymbol{\mu} - \frac{\boldsymbol{\mu}_0}{2}\right) \leq \frac{1}{4}\boldsymbol{\mu}_0^T \boldsymbol{\Omega} \boldsymbol{\mu}_0. \tag{16}$$

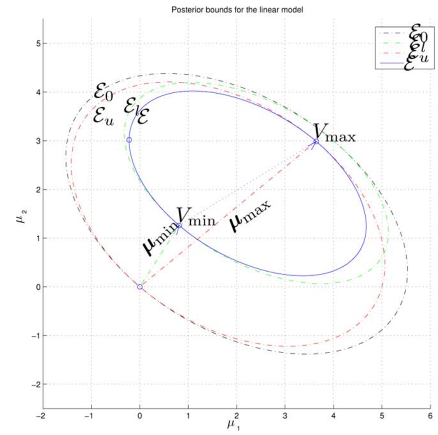

Fig. 1. Illustration of posterior bounds for a 2-D case.

- If only the lower bound constraint is imposed (i.e., $\mathbf{R}_{\boldsymbol{\theta\theta}} \succeq \mathbf{R}_{\min}$) then the bound of $\boldsymbol{\mu}$ is an ellipsoid $\mathcal{E}_l$ contained within $\mathcal{E}_0$ and $\mathcal{E}_l$ touches $\mathcal{E}_0$ at one point.
- If only the upper bound constraint is imposed (i.e., $\mathbf{R}_{\boldsymbol{\theta\theta}} \preceq \mathbf{R}_{\max}$) then the bound of $\boldsymbol{\mu}$ is an ellipsoid $\mathcal{E}_u$ (see (29)) contained within $\mathcal{E}_0$ and $\mathcal{E}_u$ touches $\mathcal{E}_0$ on its surface at the origin.
- The ellipsoid $\mathcal{E}$ of interest resides in the intersection of $\mathcal{E}_l$ (see (28)) and $\mathcal{E}_u$. It touches $\mathcal{E}_l$ at a point $V_{\min}$ and $\mathcal{E}_u$ at a point $V_{\max}$, which are, respectively, those defined by the following vectors:

$$\begin{cases} \boldsymbol{\mu}_{\min} = \boldsymbol{\Sigma}_{\min}\boldsymbol{\eta} \\ \boldsymbol{\mu}_{\max} = \boldsymbol{\Sigma}_{\max}\boldsymbol{\eta}. \end{cases} \tag{17}$$

It is noted that $\mathcal{E}$ is only a subset of $\mathcal{E}_l \cap \mathcal{E}_u$. This is due to the fact that there exists $\mathbf{R}_{\boldsymbol{\theta\theta}}$ that satisfies either the lower or upper bounds, but not both at the same time. This will be clearer subsequently.

2) Properties of the Minimax Function:

Proposition 1: For the minimax problem (8):

1) $\mathcal{S}$ is a compact and convex set;
2) $f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$ is Lipschitz continuous with respect to $\boldsymbol{\Sigma}$ and **convex** in both $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}$;
3) $g(\hat{\boldsymbol{\theta}})$ of (11) is locally Lipschitz continuous on bounded sets and also convex in $\hat{\boldsymbol{\theta}}$.

Proof: The first part is obvious due to the fact that $\mathcal{S}$ is the intersection of two translated positive semi-definite (psd) cones pointing towards each other. The second part follows from the first part and the fact that $f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$ is quadratic in both $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}$. The third part is a standard result of point-wise supremum that can be found in the convex analysis literature (for example [19]). ■

It is evident from this result that the minimax problem of interest (8) is convex for both variables $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. Hence, many computationally efficient methods developed for minimax problems, where the objective functions are concave with respect to the max problem (10), are not applicable. As shown subsequently, this max problem is confined to the comparison of extreme points in the set $\mathcal{S}$ which implies a possibly nonunique solution. Fortunately, the third part of Proposition 1 suggests that if $g(\hat{\boldsymbol{\theta}})$ can be evaluated at any $\hat{\boldsymbol{\theta}}$, then it is possible to use

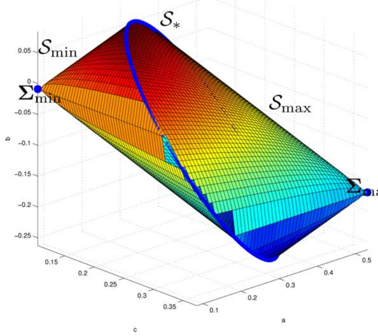

Fig. 2. Illustration of $\mathcal{S}_*, \mathbf{\Sigma}_{\min}, \mathbf{\Sigma}_{\max}$ for a 2D case. The set $\mathcal{S}_*$ is the ring where the two cones intersect.

standard unconstrained optimization routines that do not require differentiability, such as the simplex algorithm [18], to solve the min problem (11).

Note that though $f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma})$ is quadratic in both $\hat{\boldsymbol{\theta}}$ and $\mathbf{\Sigma}$, $g(\hat{\boldsymbol{\theta}})$ is more difficult to describe other than to note that it is convex, which means that second-order information is not analytically available.

One of the main issues in solving (11) is to compute $g(\hat{\boldsymbol{\theta}})$ for each $\hat{\boldsymbol{\theta}}$ and this requires the solution of the max problem (10). Note that this is a maximization of a convex function over a compact convex set, which is generally a NP-hard problem. The max problem by itself could be tackled by semi-definite relaxation to obtain an upper bound [23]. However, such an approach will not warrant the convexity of $g(\hat{\boldsymbol{\theta}})$ unless the relaxation is tight. It is therefore desirable to solve the max problem exactly. To reduce the computational cost, the geometry of the problem is exploited.

3) Solving the Minimax Problem: We first start with the following result that specifies possible candidates for problem (10).

Proposition 2: Let

$$\begin{aligned}\mathcal{S}_{\max} &\triangleq \{\mathbf{\Sigma} : \mathbf{\Sigma}_{\min} \preceq \mathbf{\Sigma} \precsim \mathbf{\Sigma}_{\max}\}\\ \mathcal{S}_{\min} &\triangleq \{\mathbf{\Sigma} : \mathbf{\Sigma}_{\max} \succeq \mathbf{\Sigma} \succsim \mathbf{\Sigma}_{\min}\}\\ \mathcal{S}_* &\triangleq \mathcal{S}_{\min} \cap \mathcal{S}_{\max} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_{\min} \precsim \mathbf{\Sigma} \precsim \mathbf{\Sigma}_{\max}\}.\end{aligned}$$

Then

$$\sup_{\mathbf{\Sigma}\in\mathcal{S}} f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}) = \max\left\{ f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}_{\min}),\ \ f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}_{\max}),\ \ \sup_{\mathbf{\Sigma}\in\mathcal{S}_*} f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}) \right\}. \tag{18}$$

Proof: It is easy to verify from the definition of the boundary of a set that $\mathrm{bd}(\mathcal{S}) = \mathcal{S}_{\min} \cup \mathcal{S}_{\max}$. In other words, $\mathcal{S}_{\min}$ and $\mathcal{S}_{\max}$ are the boundaries of the two translated psd cones corresponding to the upper and lower bounds on $\mathbf{\Sigma}$. Further, it is possible to verify that any $\mathbf{\Sigma} \in \mathcal{S}_{\min}$ is a convex combination of $\mathbf{\Sigma}_{\min}$ and a point in $\mathcal{S}_*$. Similarly, any point $\mathbf{\Sigma} \in \mathcal{S}_{\max}$ is a convex combination of $\mathbf{\Sigma}_{\max}$ and a point in $\mathcal{S}_*$. Together with the convexity of $f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma})$ with respect to $\mathbf{\Sigma}$ over $\mathcal{S}$, we can conclude that the supremum is found at $\mathcal{S}_* \cup \{\mathbf{\Sigma}_{\min}, \mathbf{\Sigma}_{\max}\}$ (i.e., they are extreme points of the set $\mathcal{S}$, see Fig. 2 for an illustration). ■

For notational convenience, we introduce

$$\begin{aligned}\boldsymbol{g}_{\max}(\hat{\boldsymbol{\theta}}) &\triangleq f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}_{\max})\\ \boldsymbol{g}_{\min}(\hat{\boldsymbol{\theta}}) &\triangleq f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}_{\min})\\ \boldsymbol{g}_{*}(\hat{\boldsymbol{\theta}}) &\triangleq \sup_{\mathbf{\Sigma}\in\mathcal{S}_*} f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}).\end{aligned}$$

We shall remove $\boldsymbol{g}_{\min}$ from the list of the possible candidates for the max problem. Before doing so, we study the geometry of the set $\mathcal{S}_*$ via the following result.

Proposition 3: If the set $\mathcal{S}_*$ is not empty, there exists a nonempty subset $\mathcal{S}_*^b \subseteq \mathcal{S}_*$ such that the mapping from $\mathcal{S}_*^b$ to the boundary of $\mathcal{E}$

$$\mathcal{S}_*^b \longrightarrow bd(\mathcal{E}) : \boldsymbol{\mu} = \mathbf{\Sigma}\boldsymbol{\eta}$$

is *bijective*, i.e., each $\mathbf{\Sigma} \in \mathcal{S}_*^b$ maps to some $\boldsymbol{\mu} = \mathbf{\Sigma}\boldsymbol{\eta} \in \mathrm{bd}(\mathcal{E})$ and for each $\boldsymbol{\mu} \in \mathrm{bd}(\mathcal{E})$ there exists a unique $\mathbf{\Sigma} \in \mathcal{S}_*^b$ such that $\boldsymbol{\mu} = \mathbf{\Sigma}\boldsymbol{\eta}$. Further, we have $\mathcal{S}_*^b \equiv \mathcal{S}_*$ for the case $K = 2$. The proof is detailed in Appendix E.

By applying the result of Proposition 3, we can remove $g_{\min}(\hat{\boldsymbol{\theta}})$ from the list and simplify the original minimax problem as follows.

Corollary 1: When $K > 1$ the set $\mathcal{S}_*$ is not empty and:

$$g_*(\hat{\boldsymbol{\theta}}) \geq g_{\min}(\hat{\boldsymbol{\theta}})$$

so that the minimax problem (8) is equivalent to

$$\inf_{\hat{\boldsymbol{\theta}}\in\mathcal{E}} \max\left\{ g_{\max}(\hat{\boldsymbol{\theta}}),\ g_*(\hat{\boldsymbol{\theta}}) \right\}. \tag{19}$$

The proof is detailed in Appendix F.

Define $\gamma = \boldsymbol{\eta}^T(\mathbf{\Sigma}_{\max} - \mathbf{\Sigma}_{\min})\boldsymbol{\eta}$. It naturally follows that the minimax solution depends on the relative geometry of $g_*(\hat{\boldsymbol{\theta}})$ and $g_{\max}(\hat{\boldsymbol{\theta}})$ as follows:

Theorem 1:

- Case 1: If $\gamma \leq 1$ which implies $g_{\max}(\boldsymbol{\mu}_{\max}) \geq g_*(\boldsymbol{\mu}_{\max})$, the minimax solution of (19) is

$$\hat{\boldsymbol{\theta}}^* = \boldsymbol{\mu}_{\max}.$$

- Case 2: If $\gamma > 1$ then:

$$\hat{\boldsymbol{\theta}}^* = \arg\min_{\hat{\boldsymbol{\theta}}} \max\left\{ g_*(\hat{\boldsymbol{\theta}}), g_{\max}(\hat{\boldsymbol{\theta}}) \right\}.$$

The proof is detailed in Appendix G.

We note that the condition $\gamma \leq 1$ was proved for a special case where $\mathbf{C} = \mathbf{I}$ and $\mathbf{\Sigma}_{\max} - \mathbf{\Sigma}_{\min} = \lambda\mathbf{I}$ in [9]. In other words, the ellipsoid $\mathcal{E}$ was assumed to be spherical. Here, we do not need to make such a restrictive assumption. The two cases specified in Theorem 1 are illustrated, respectively, in Figs. 3 and 4. In these figures the functions $g_{\max}(\boldsymbol{\theta})$ and $g_*(\boldsymbol{\theta})$ are plotted as $\boldsymbol{\theta}$ varies within the ellipsoid. It is also of interest to compare the minimax solution with an obvious choice corresponding to the average of the upper and lower bounds, which is the center of the ellipsoid. In Fig. 3, $g_{\max}(\boldsymbol{\theta})$ is always greater than $g_*(\boldsymbol{\theta})$ over the ellipsoid $\mathcal{E}$, resulting in the minimax solution being the minimum of $g_{\max}(\boldsymbol{\theta})$, which is $\boldsymbol{\mu}_{\max}$. Also, if one selects

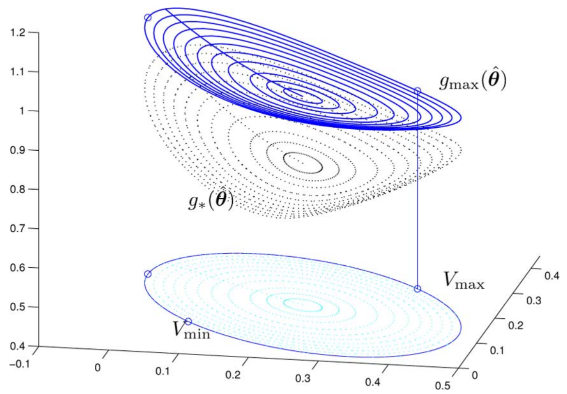

Fig. 3. Case 1. $\boldsymbol{\eta}^T(\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta} \leq 1$ which implies $g_{\max}(\boldsymbol{\mu}_{\max}) \geq g_*(\boldsymbol{\mu}_{\max})$.

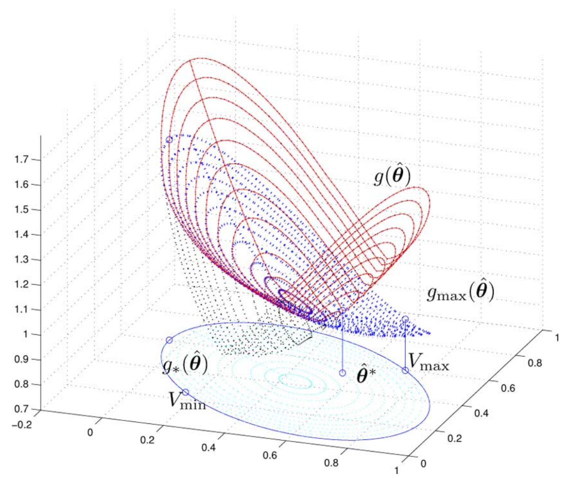

Fig. 4. Case 2: $\boldsymbol{\eta}^T(\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta} > 1$.

the average of the upper and lower bounds, which corresponds to the center of the ellipsoid, then the maximum risk associated with this choice is larger than the minimax risk given by $g_{\max}((\boldsymbol{\mu}_{\max} + \boldsymbol{\mu}_{\min})/2) - g_{\max}(\boldsymbol{\mu}_{\max})$. In Fig. 4, the minimax solution is found as the minimum of the intersection between $g_{\max}(\hat{\boldsymbol{\theta}})$ and $g_*(\hat{\boldsymbol{\theta}})$. In this particular setting, the minimax solution is away from the center of the ellipsoid $\mathcal{E}$, meaning that the maximum risk associated with the average choice is significantly greater than the minimax risk.

The difficult part of the problem is to derive the minimax solution when $\gamma > 1$. Before deriving such a solution, we note an important characteristic of the minimax solution as γ varies in the range $(0, \infty)$. When $0 < \gamma \leq 1$, Theorem 1 shows that the minimax always stays at $\boldsymbol{\mu}_{\max}$. A closer look reveals that the minimax objective function (9) consists of two terms. In this case the supremum of the second term $\mathrm{tr}[\boldsymbol{\Sigma}]$ is larger than the supremum of the first term due to Case 1 of Theorem 1, so that the choice $\boldsymbol{\mu}_{\max}$ is obvious. When $\gamma \to \infty$, the supremum of the second term is negligible compared with the supremum of the first term and the minimax objective function can be approximated by removing the second term. It can then be easily seen that the minimax solution when $\gamma \to \infty$ is the center of the ellipsoid $\mathcal{E}$ (i.e., $\boldsymbol{\mu}_c$). Intuitively, we would expect that the minimax solution is continuous in $\mathcal{E}$ as γ varies. Thus, for $\gamma > 1$, the minimax solution must follow some trajectory from $\boldsymbol{\mu}_{\max}$ to $\boldsymbol{\mu}_c$. When the ellipsoid $\mathcal{E}$ is not skewed, the geometrical symmetry hints that the minimax solution, whilst being a function of γ, must also depend on some convex combination of $\boldsymbol{\mu}_c$ and $\boldsymbol{\mu}_{\max}$ such that it becomes $\boldsymbol{\mu}_{\max}$ when $\gamma \to 1$ and $\boldsymbol{\mu}_c$ when $\gamma \to \infty$. One conjecture that reflects this behavior when $\gamma > 1$ is

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mu}_c + \frac{(\boldsymbol{\mu}_{\max} - \boldsymbol{\mu}_c)}{\gamma}. \tag{20}$$

For a very special case when the ellipsoid $\mathcal{E}$ becomes a sphere, DasGupta and Studden [9] have achieved the following result.

Theorem 2: For problem (8), with $\mathbf{C} = \mathbf{I}$ and $\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min} = \lambda\mathbf{I}$, (20) is conditional Γ-minimax.

It is of interest to note that DasGupta and Studden start from a study of the Γ-minimax solution for the linear combination $\mathbf{c}^T\boldsymbol{\theta}$ and achieve the explicit minimax solution in the same form. However, as the mapping from $\boldsymbol{\theta}$ to $\mathbf{c}^T\boldsymbol{\theta}$ is *not bijective*, they were unable to derive the minimax solution for the general case, except for the special case as detailed above.

In what follows, we show that (20) is the conditional Γ-minimax for a slightly more general case when the ellipsoid $\mathcal{E}$ is not necessarily spherical. We first derive the following result as a consequence of Theorem 1.

Proposition 4: For the case $\gamma > 1$, there exist nonempty subsets $\mathcal{E}_1$ and $\mathcal{E}_2$ ($\mathcal{E}_1 \cup \mathcal{E}_2 = \mathcal{E}$) such that

$$\begin{cases} g_{\max}(\hat{\boldsymbol{\theta}}) \leq g_*(\hat{\boldsymbol{\theta}}) & \forall\, \hat{\boldsymbol{\theta}} \in \mathcal{E}_1 \\ g_{\max}(\hat{\boldsymbol{\theta}}) \geq g_*(\hat{\boldsymbol{\theta}}) & \forall\, \hat{\boldsymbol{\theta}} \in \mathcal{E}_2. \end{cases}$$

The minimax solution lies in the intersection of $\mathcal{E}_1$ and $\mathcal{E}_2$, where $g_{\max}(\hat{\boldsymbol{\theta}}) = g_*(\hat{\boldsymbol{\theta}}), \forall \hat{\boldsymbol{\theta}} \in \mathcal{E}_1 \cap \mathcal{E}_2$.

The proof of this result is detailed in Appendix H.

Next, we note an important observation that if $g(\hat{\boldsymbol{\theta}}_*)$ is a local minimum for a small neighborhood about some $\hat{\boldsymbol{\theta}}_*$, then the convexity of $g(\hat{\boldsymbol{\theta}})$ implies that $\hat{\boldsymbol{\theta}}_*$ must be the minimax solution. Using the intuitive result proposed by DasGupta and Studden [9], we formally state the following sufficient condition for the minimax solution:

Proposition 5: For the case $\gamma > 1$, if there exist $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}_* \in \mathcal{S}_*$ such that:

- $\hat{\boldsymbol{\theta}}$ is some convex combination of $\boldsymbol{\mu}_{\max}$ and $\boldsymbol{\Sigma}_*\boldsymbol{\eta}$;
- $f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\max}) = \sup_{\boldsymbol{\Sigma} \in \mathcal{S}_*} f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) = f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_*)$;

then $\hat{\boldsymbol{\theta}}$ is the minimax solution.

The proof is detailed in Appendix I.

Proposition 5 leads to the following result that slightly generalizes Theorem 2:

Theorem 3: Denote $\mathbf{U}$ and $\boldsymbol{\Lambda}$ as the eigenvectors and eigenvalue matrix of $\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min}$ so that $\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$. Suppose that the top K eigenvalues are the same and denote the corresponding eigenvectors by $\mathbf{U}_K$. Then (20) is the conditional Γ-minimax solution if $\boldsymbol{\eta} \in \mathrm{span}\{\mathbf{U}_K\}$.

The proof is detailed in Appendix J.

Geometrically speaking, Theorem 3 asserts that (20) is the minimax solution only if $\boldsymbol{\eta}$ lies in the convex hull of the norm vectors of the top K principal axes of ellipsoid $\mathcal{E}$ whose lengths are equal. It can be easily seen that the condition imposed in DasGupta and Studden is a special case where all eigenvalues of $\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min}$ are the same so that $\mathrm{span}\{\mathbf{U}_K\}$ is actually the whole space. Thus, Theorem 3 would be useful in practice if the number of principal and identical eigenvalues is large or when there is a large number of principal eigenvalues which are close to each other.

If the condition in Theorem 3 is not met, practical experience shows that the minimax solution (20) is unlikely to be a convex combination of $\boldsymbol{\mu}_{\max}$ and $\boldsymbol{\mu}_c$. It appears from the proof of Theorem 3 that the distance between (20) and the true minimax solution depends both on the flatness of the ellipsoid $\mathcal{E}$ (i.e., the ratio between the maximum and minimum eigenvalues of $\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min}$) and the value of $\boldsymbol{\eta}$. Accordingly, we propose the following algorithm to compute a more accurate approximation of the minimax estimate for low-dimensional problems when $\gamma > 1$.

B. Algorithm

To numerically compute the minimax solution, one possible approach is to calculate $g(\hat{\boldsymbol{\theta}})$ and exploit its convexity. This necessitates the evaluation of $g_*(\hat{\boldsymbol{\theta}})$ as shown in (19). As mentioned, (10) is a difficult optimization problem except for some special cases. We can also view $g_*(\hat{\boldsymbol{\theta}})$ as a semi-infinite max problem over $\mathcal{S}_*$ (i.e., the maximum[3] over an *infinite* number of suitably chosen points $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots \in \mathcal{S}_*$ (see also [19]):

$$g_*(\hat{\boldsymbol{\theta}}) = \max_{i=1,2,\ldots} \left\{ f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_i) \right\}. \tag{21}$$

This suggests that if it is possible to obtain a dense sampling of the parameter space $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots \in \mathcal{S}_*$ then $g(\hat{\boldsymbol{\theta}})$ can be well approximated. The question then is how to select the samples $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots \in \mathcal{S}_*$. Proposition 3 suggests that at least for the case $K = 2$, every $\boldsymbol{\Sigma} \in \mathcal{S}_*$ maps to a boundary point of the ellipsoid $\mathcal{E}$, and thus sampling on the boundary of $\mathcal{E}$ can provide an equivalent sampling on $\mathcal{S}_*$ via a bijective mapping. Even though we have not yet obtained a proof for a general case, intensive numerical studies lead to a conjecture that this may hold for $K = 3$ as well (see Appendix M for an example). A rigorous proof for the case $K \geq 3$ is an unsolved problem. For the case $K > 3$, the following is proposed. Denote a boundary point on ellipsoid $\mathcal{E}$ as $\boldsymbol{\mu}_l$. It follows that the associated $\boldsymbol{\Sigma}_l \in \mathcal{S}_*$ can only be a possible candidate for the max problem when it is a solution of

$$\begin{aligned} \max \quad & \mathsf{tr}[\boldsymbol{\Sigma}] \\ \text{s.t} \quad & \boldsymbol{\Sigma}\boldsymbol{\eta} = \boldsymbol{\mu}_l,\ \boldsymbol{\Sigma}_{\min} \preceq \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_{\max}. \end{aligned} \tag{22}$$

Once a set of $\boldsymbol{\mu}_l$ and $\boldsymbol{\Sigma}_l$, $l = 1, \ldots, L$, has been computed, the evaluation of $g_*(\hat{\boldsymbol{\theta}})$ is straightforward

$$g_*(\hat{\boldsymbol{\theta}}) = \max_{l \in \{1,2,\ldots,L\}} \left\{ \| \hat{\boldsymbol{\theta}} - \boldsymbol{\mu}_l \|^2 + \mathsf{tr}[\boldsymbol{\Sigma}_l] \right\}.$$

The minimax solution is then found from the following problem:

$$\hat{\boldsymbol{\theta}}^* = \arg\min_{\hat{\boldsymbol{\theta}}} \max \left\{ g_*(\hat{\boldsymbol{\theta}}), g_{\max}(\hat{\boldsymbol{\theta}}) \right\}$$

using a nondifferentiable unconstrained optimization technique with the starting point at (20).

Next, we detail two important aspects of the proposed algorithm.

[3]Technically, $\mathcal{S}_*$ contains an uncountable, infinite number of points. However, for practical purposes considering a countable, but dense, set of points does not change the result.

- Selection of L points on the boundary of ellipsoid $\mathcal{E}$: for each pair of eigenvectors $\mathbf{u}_i, \mathbf{u}_j, i \neq j$ of $\frac{4\boldsymbol{\Phi}}{\gamma}$ (cf. (14)), the boundary points can be selected as
$$\begin{aligned} \boldsymbol{\mu}_l &= \boldsymbol{\mu}_c + \left[\frac{\mathbf{u}_i}{\sqrt{\lambda_i}} \frac{\mathbf{u}_j}{\sqrt{\lambda_j}} \right] \begin{bmatrix} \cos(\alpha_l) \\ \sin(\alpha_l) \end{bmatrix}, \\ \alpha_l &= 2\pi l / L_0, l = 1, \ldots, L_0 \end{aligned}$$
where L_0 is the number of selected points for each pair $\mathbf{u}_i$, $\mathbf{u}_j$, which means $L = K(K+1)L_0/2$. The above sampling is representative of the full search space for $K = 2$. For $K \geq 3$, one can make the selection denser by considering multiple eigenvectors at the same time and using multiple weights whose sum-of-squares equals 1 rather than the weights defined by $\cos(\alpha)$ and $\sin(\alpha)$ as above.
- The problem (22) can be easily converted to a standard semi-definite programming (SDP) formulation that includes a linear objective function, equality constraints, and semi-definite constraints on the matrix variables [1], [15], [23]. Once an optimization problem is formulated in an SDP framework, it is generally considered *solved*, at least numerically [15], for suitable problem size and computing power (see Appendix K for an example).

Remarks:

- From Proposition 3, it is clear that when $K = 2$, our proposed algorithm is numerically exact, in the sense that the error of the minimax solution can be made arbitrarily small by selecting sufficiently large number of points on the boundary of the ellipsoid $\mathcal{E}$. Although, the case $K > 3$ is still an open question, the proposed algorithm may still be useful in potential applications where the number of unknown parameters is small, such as 2-D or 3-D tracking with prior dynamic uncertainty in computer vision.
- In addition, empirical investigation shows that the approximation is generally satisfactory for a relatively small value of L. For example, when $L = 12$ and $K = 2$, the relative error $\frac{\|\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}_L^*\|^2}{\|\boldsymbol{\mu}_{\max} - \boldsymbol{\mu}_{\min}\|^2}$ is of the order of less than 0.5%.
- From Theorem 1, there are two possible cases. These arise from the structure of the conditional risk function (9), which is the objective function of the minimax problem. It consists of two terms, the first measures the distance of the estimate to a posterior mean, the second measures the trace of $\boldsymbol{\Sigma}$. The minimax problem is essentially the compromise of these two terms over the ellipsoid $\mathcal{E}$. For tractability, we consider the special case with $\mathbf{R}_{\boldsymbol{\theta\theta}} = \sigma^2\mathbf{I}$, $\mathbf{R}_{\mathbf{nn}} = \sigma_0^2\mathbf{I}$, and $\mathbf{C} = \mathbf{I}$. Then it is possible to show that
$$\begin{aligned} \boldsymbol{\Sigma} &= (\sigma^{-2} + \sigma_0^{-2})^{-1}\mathbf{I} \\ \boldsymbol{\eta} &= \sigma_0^{-1}\mathbf{u} \end{aligned}$$
where $\mathbf{u}$ is a random variable drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I})$. It is straightforward to show that
$$\frac{\mathsf{tr}[\boldsymbol{\Sigma}]}{\| \boldsymbol{\Sigma}\boldsymbol{\eta} \|^2} = O\left(K(1 + \sigma_0^2/\sigma^2)\right).$$
A closer look reveals that when the measurement noise $\mathbf{n}$ is *"large"* (relative to the signal) (i.e., $\sigma_0^2 \gg \sigma^2$) the second term of the conditional risk will become dominant. In other

words, the difference between $\text{tr}[\mathbf{\Sigma}_{\max}]$ and $\sup_{\mathbf{\Sigma}\in\mathcal{S}_*}\text{tr}[\mathbf{\Sigma}]$ is much greater than the size of any principal axis of the ellipsoid $\mathcal{E}$. This is Case 1 of Theorem 1, where the minimax solution is found at $\boldsymbol{\mu}_{\max}$, and no complicated computations need to be carried out. When the noise is sufficiently "*small*" or *"moderate"*, these two terms could be comparable. Hence, one needs to invoke the proposed algorithm to solve Case 2 of Theorem 1 if the conditions in Theorem 3 are not satisfied.

III. Experimental Results

In this section, we numerically demonstrate the proposed method to address two important questions that arise. The first question is about the accuracy of the proposed algorithm in obtaining the conditional Γ-minimax solution. The solution by DasGupta and Studden, which is the only tractable solution known in the literature and only applicable for a special case, is also considered. The second question is the performance of the conditional Γ-minimax formulation relative to other Bayesian formulations in terms of more commonly used measures such as mean squared error (MSE). While such a comparison to other Bayesian formulations has been considered previously in, for example, [9], a new comparison is provided to reflect the new setting. We note importantly that while the analytical solution in Theorem 3 is valid for all K, the algorithm is only numerically exact for $K \leq 2$. Thus, we only perform the detailed analysis for $K = 2$.

The parameters are selected somewhat arbitrarily as follows (Matlab format): the design matrix $\mathbf{C} = [3, 0.5; -0.1, 1]$, the noise covariance matrix $\mathbf{R_{nn}} = \sigma^2[4, 0.5; 0.5, 1], \sigma^2 = 1$, the bounds are $\mathbf{R}_{\min} = 0.1[4, -0.4; -0.4, 1]$ and $\mathbf{R}_{\max} = 2[4, 0.4; 0.4, 1]$, and $\boldsymbol{\mu_\theta} = \mathbf{0}$.

We parameterize the true covariance matrix $\mathbf{R}_{\boldsymbol{\theta\theta}}$ as

$$\mathbf{R}_{\boldsymbol{\theta\theta}} = \mathbf{R}_{\min} + \sum_{i=1}^{2} \rho_i \lambda_i \mathbf{u}_i \mathbf{u}_i^T \tag{23}$$

where $\mathbf{u}_i$'s are the eigenvectors of $\mathbf{R}_{\max} - \mathbf{R}_{\min}$ associated with the corresponding eigenvalues λ_i, $i = 1, 2$ and $\rho_i \in [0, 1]$. We sample the true covariance matrix $\mathbf{R}_{\boldsymbol{\theta\theta}}$ uniformly in terms of ρ_i, the sampling points are shown in Fig. 5. In particular, the sampling point 1 corresponds to $\mathbf{R}_{\min}$ and the sampling point 25 corresponds to $\mathbf{R}_{\max}$. This choice ensures that the selected covariance matrix satisfies $\mathbf{R}_{\min} \preceq \mathbf{R}_{\boldsymbol{\theta\theta}} \preceq \mathbf{R}_{\max}$.

In all experiments, we randomly generate a number of $\boldsymbol{\theta}$'s for each selected covariance matrix $\mathbf{R}_{\boldsymbol{\theta\theta}}$. Then, we randomly generate a number of $\mathbf{y}$'s according to the normal model (1) for each randomly generated $\boldsymbol{\theta}$. We then report the average performance over both $\boldsymbol{\theta}$ and $\mathbf{y}$.

A. Accuracy of the Proposed Algorithm

In the first experiment, we compare the accuracy of the proposed algorithm with respect to the number of sampling points L on the boundary of the ellipsoid $\mathcal{E}$ as mentioned previously. Specifically, we select $L = 12$, 20, 44 and measure the average squared precision error relative to the true minimax solution defined as $||\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{\text{mnm}}||_2$, where $\hat{\boldsymbol{\theta}}$ is the minimax solution obtained from one of the methods being compared, and $\hat{\boldsymbol{\theta}}_{\text{mnm}}$ is the true

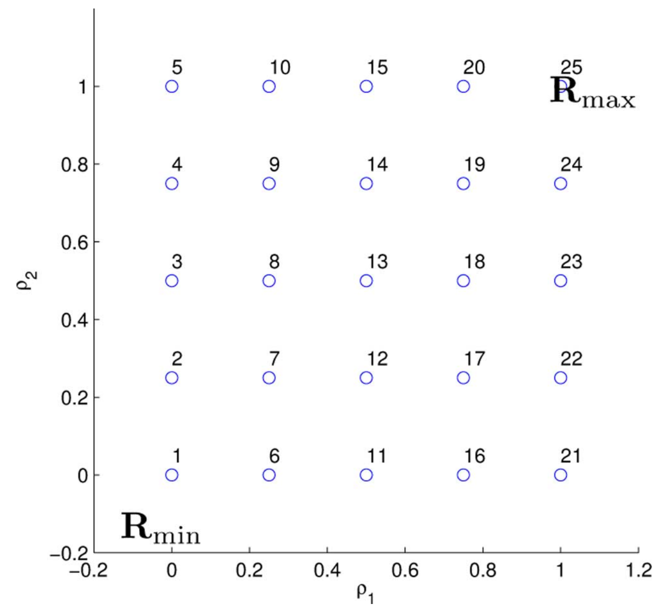

Fig. 5. Sampling points for the prior covariance matrix $\mathbf{R}_{\boldsymbol{\theta\theta}}$ used in the experiment.

minimax solution. This true minimax solution is numerically obtained by setting L to be very large. We also compare such errors with the error induced when using the solution by DasGupta and Studden for the general case. This gives an idea of how much accuracy is gained when using our algorithm. When running the experiment, we observe that when $\mathbf{R}_{\boldsymbol{\theta\theta}}$ varies as above, on average, the solution is found at $\boldsymbol{\mu}_{\max}$ ($\gamma > 1$) for 20% of the time. In this case, a direct application of DasGupta and Studden's result yields the exact solution. For the case when the minimax solution is not found at $\boldsymbol{\mu}_{\max}$, Fig. 6 plots the errors of the estimates. Here, the errors are plotted against the sampling points defined graphically in Fig. 5. We observe that when $\mathbf{R}_{\boldsymbol{\theta\theta}}$ varies, the accuracy of all methods varies slightly, possibly due to numerical properties. However, the overall trend is that when the number of sampling points L is large, the proposed minimax algorithm achieves better accuracy and can be ten times more accurate than a direct application of DasGupta and Studden's solution.

B. MSE Performance

The Γ-minimax solution is formulated to cope with the worst case scenario of conditional risk. As mentioned by DasGupta and Studden [9], it should be examined using other conventional criteria such as MSE (with respect to the true value of $\boldsymbol{\theta}$ as mentioned previously). This is what we demonstrate next. The Γ-minimax solution is compared with direct implementations of the following methods.

- Parametric empirical Bayes. We extend the solution [7, p. 76] to the multivariate case, with a minor change: the maximum likelihood estimate of the covariance matrix, which is singular for the problem being considered here, is replaced by the shrinkage method recently described in [20], wherein the shrinkage density is automatically found. The parametric empirical Bayes is

$$\hat{\boldsymbol{\theta}}_{\text{PEB}} = \boldsymbol{\mu_\theta} + \mathbf{C}^{-1}(\mathbf{\Sigma_{yy}} - \mathbf{R_{nn}})^{+}\mathbf{\Sigma_{yy}}^{-1}(\mathbf{y} - \mathbf{C}\boldsymbol{\mu_\theta})$$

where $()^+$ is the projection to the positive semi-definite cone, and $\boldsymbol{\Sigma}_{\mathbf{yy}}$ is simply the shrinkage estimate based on $\mathbf{y}$.

- *Hierarchical Bayes.* We consider the covariance matrix $\mathbf{R}$ of $\boldsymbol{\theta}$ as a random variable and obtain the estimate as

$$\hat{\boldsymbol{\theta}}_{\mathrm{HB}} = \frac{1}{\int_{\boldsymbol{\theta}} f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}} \int_{\boldsymbol{\theta}} \boldsymbol{\theta} f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{24}$$

where

$$f(\boldsymbol{\theta}) = \int_{\mathbf{R}} f(\boldsymbol{\theta}|\mathbf{R}) f(\mathbf{R}) d\mathbf{R}.$$

Using the popular inverse Wishart for the hyperprior $\pi(\mathbf{R})$ leads to high computational cost. Hence, we consider the choice where $\mathbf{R} = \tau^2\mathbf{I}$, where τ^2 follows the χ^2_m distribution. This enables the integration over the continuous random variable τ to be easier.

- The minimax estimates by Eldar and Merhav proposed in [10]. To adapt their methods to our settings, we simply set the eigenvalues of $\boldsymbol{\Sigma}_{\max}$ and $\boldsymbol{\Sigma}_{\min}$ as the bounds on the eigenvalues of $\boldsymbol{\Sigma}$ for the first model, and the maximum singular value of $\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min}$ for the second model.
- The optimal Bayes solution. This solution is obtained when $\mathbf{R}_{\boldsymbol{\theta\theta}}$ is exactly known and the estimate is taken as the posterior mean as shown in (4). Apparently, this solution serves as the lower bound on the MSE of all compared methods (i.e., the optimal Bayes solution is exactly the minimum MSE (MMSE) solution).

The MSE is defined as the average of $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2$ where $\hat{\boldsymbol{\theta}}$ is the estimate from one of the methods being compared and $\boldsymbol{\theta}$ is the true value of the random parameters. The results on MSE performance are shown in Fig. 7. As expected, the optimal Bayes solution achieves the overall minimum MSE. We also note that there seem to be "favorable" values for the covariance matrix $\mathbf{R}_{\boldsymbol{\theta\theta}}$ at which all algorithms appear to achieve better accuracy than others at sampling indices 1, 6, 11, 16, 21. By referring back to Fig. 5, they correspond to the cases where $\rho_2 = 0$ (cf. (23)). At these points, the Γ-solution appears to achieve a slightly better accuracy than all other compared methods, and this is close to the optimal Bayes solution. Overall, the Γ-minimax solution appears to have similar MSE performance as the parametric Bayes or the Eldar-Merhav minimax estimates, though its variability is slightly larger. We note importantly that Eldar-Merhav minimax estimates directly minimize the maximum MSE whilst the Γ-minimax solution only optimizes the conditional risk.

We also note that our experiment considers some particular implementations of parametric empirical Bayes and hierarchical Bayes, and no conclusion about these two general approaches is made here. Rather, these particular implementations help us to better understand the robustness and accuracy of the proposed algorithm in solving the minimax solution. Although the proposed algorithm is theoretically justified in terms of minimax properties in the conditional Bayesian framework, it also demonstrates good MSE performance. Our findings are consistent with [9] conducted in a different setup.

IV. Conclusion

We addressed the robustness issue with respect to prior uncertainty in the multivariate normal model. This problem is practically important for Bayesian data analysis where one cannot obtain a complete prior specification. A theoretical study on the minimax solution was conducted under the Bayesian conditional Γ-minimax approach. An extension to a previous result by DasGupta and Studden, when the ellipsoidal bound of the conditional mean is not spherical, was obtained. For the more general case, we proposed a practical algorithm to compute the minimax estimate. This algorithm is numerically exact for the case $K \leq 2$ and is based on a geometrical interpretation of the space of the posterior mean, and by exploiting the problem's geometry. We also discussed technical issues related to implementation. Numerical studies show that the proposed algorithm improves minimax solution accuracy in the general case and achieves robustness as well as desirable MSE properties.

Appendix A
Discussion on a Related Work

In what follows, we give a discussion on related minimax settings by Eldar and Merhav [10] via both theoretical analysis and a numerical example.

1) Simplified Minimax Solution: First, we show that the conditional Γ-minimax problem can be readily solved using the results presented in this paper when the constraint set $\mathcal{S}$ is relaxed as in [10]. This discussion illustrates two important arguments.

- The positive semi-definite constraints are already sufficiently general.
- When *simpler* constraints, which are subsets of the constraints considered in this work, are imposed, such as those discussed in the work of Eldar and Merhav [10], one can readily derive an explicit minimax estimator under the conditional Γ-minimax paradigm.

First, we recall the simplified form of the minimax problem of interest

$$\arg\min_{\boldsymbol{\theta}} \max_{\boldsymbol{\Sigma}\in\mathcal{S}} \|\boldsymbol{\theta} - \boldsymbol{\Sigma\eta}\|_2^2 + \mathrm{tr}[\boldsymbol{\Sigma}]$$

where $\mathcal{S} = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma}_{\min} \preceq \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_{\max}\}$. It is noted that:

- psd constraints imply determinant constraints but not the converse;
- psd constraints imply trace constraints but not the converse;
- psd constraints imply norm constraints but not the converse.

Considering the psd constraints also makes the problem difficult to solve. However, we have obtained analytical results under some special cases as well as a practical algorithm for other cases.

Now, consider the two constraints in Eldar and Merhav [10].

- Type 1: $\mathcal{S}_1 = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma} \text{ diagonal}, \Sigma_{\min,ii} \leq \Sigma_{ii} \leq \Sigma_{\max,ii}\}$.
- Type 2: $\mathcal{S}_2 = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c + \boldsymbol{\Delta}, \|\boldsymbol{\Delta}\| \leq \varepsilon\}$.

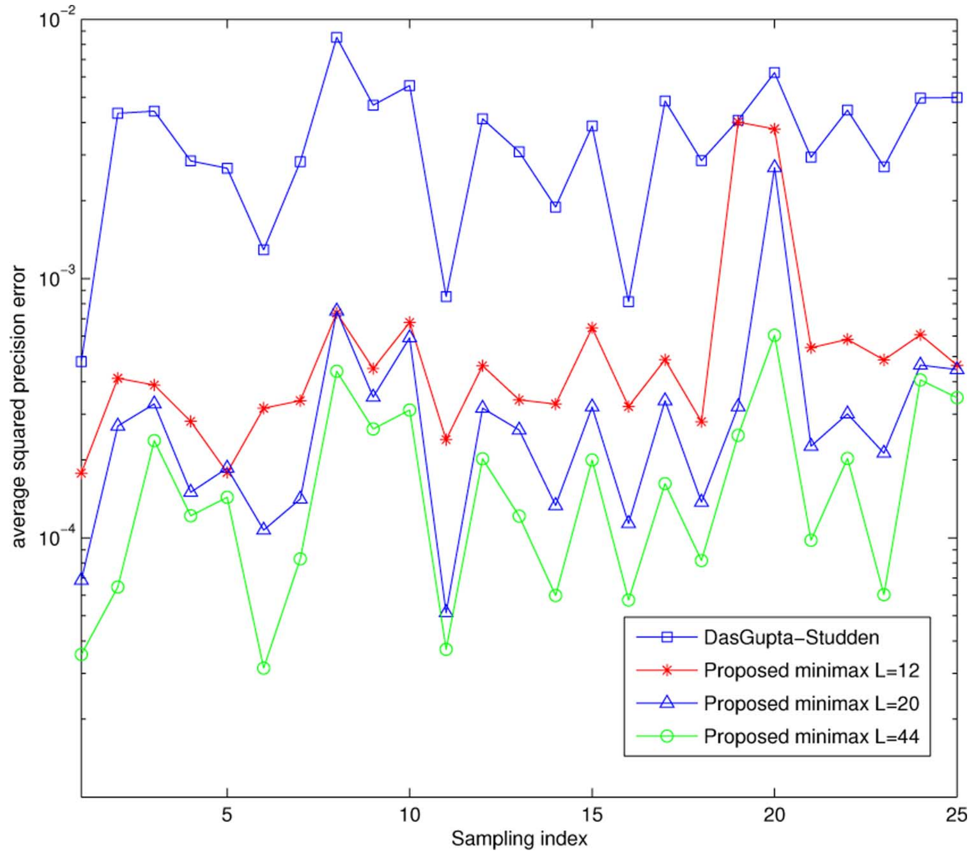

Fig. 6. Average squared precision error relative to the true minimax solution.

We have two remarks regarding our adaptation of Eldar and Merhav's settings.

- For $\mathcal{S}_1$: Eldar and Merhav consider the case where, in our own notation, $\mathbf{R}_{\boldsymbol{\theta}}$ and $\mathbf{\Omega}$ share the same eigenvectors, and that the eigenvalues of $\mathbf{R}_{\boldsymbol{\theta\theta}}$ are upper and lower bounded. It can be easily shown that $\mathbf{\Sigma}$ also shares the same eigenvectors and its eigenvalues are also upper and lower bounded. As the minimax problem of interest is invariant to an unitary transformation, we can consider $\mathbf{\Sigma}$ being diagonal without loss of generality.
- For $\mathcal{S}_2$: The eigenvalues of $\mathbf{\Delta}$ are bounded between $\pm\varepsilon$.

Consider the set $\mathcal{S}_1$ as a relaxing condition for the set $\mathcal{S}$. In this case, $\mathbf{\Sigma}$ is only diagonal. Thus, we can write the objective function of the minimax problem as follows:

$$f(\boldsymbol{\theta},\mathbf{\Sigma}) = ||\boldsymbol{\theta} - \mathbf{\Sigma}\boldsymbol{\eta}||_2^2 + \mathrm{tr}[\mathbf{\Sigma}] = \sum_{k=1}^{K}(\theta_k - \Sigma_{kk}\eta_k)^2 + \Sigma_{kk}.$$

Thus, we can *decouple* the original K-dimensional minimax problem to K independent *univariate* minimax problems

$$\min_{\theta_k}\max_{\Sigma_{kk}}\sum_{k=1}^{K}(\theta_k - \Sigma_{kk}\eta_k)^2 + \Sigma_{kk} = \sum_{k=1}^{K}\min_{\theta_k}\max_{\Sigma_{kk}}(\theta_k - \Sigma_{kk}\eta_k)^2 + \Sigma_{kk}.$$

The solution of each univariate minimax subproblem can be found via the trivial result for the special case $K = 1$. The result by DasGupta and Studden is then applicable.

Next, consider the set $\mathcal{S}_2$. We note that ε needs to be chosen properly so that $\mathbf{\Sigma} \succ \mathbf{0}$ and $\mathbf{\Delta}$ does not have to be diagonal. As the eigenvalues of $\mathbf{\Delta}$ are bounded between $\pm\varepsilon$, its follows that $-\varepsilon\mathbf{I} \preceq \mathbf{\Delta} \preceq \varepsilon\mathbf{I}$. We can set $\mathbf{\Sigma}_{\min} = \mathbf{\Sigma}_c - \varepsilon\mathbf{I}$ and $\mathbf{\Sigma}_{\max} = \mathbf{\Sigma}_c + \varepsilon\mathbf{I}$ so that $\mathbf{\Sigma}_{\min} \preceq \mathbf{\Sigma} \preceq \mathbf{\Sigma}_{\max}$. As the ellipsoid becomes the sphere in this case, the result of DasGupta and Studden directly applies.

In conclusion, when a simpler constraint set is imposed, conditional Γ-minimax estimates can be analytically derived from the problem considered in the manuscript.

2) Numerical Example: Next, we numerically demonstrate the choices between different strategies.

- The true conditional Γ-minimax estimate $\hat{\boldsymbol{\theta}}$
- The DasGupta-Studden estimate $\hat{\boldsymbol{\theta}}_{\mathrm{DGS}}$
- The Eldar-Merhav estimate for constraint $\mathcal{S}_1$, $\hat{\boldsymbol{\theta}}_{\mathrm{EM1}}$
- The Eldar-Merhav estimate for constraint $\mathcal{S}_2$, $\hat{\boldsymbol{\theta}}_{\mathrm{EM2}}$

We consider the following settings:

$$\mathbf{\Sigma}_{\min} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{\Sigma}_{\max} = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}, \boldsymbol{\eta} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \mathbf{\Omega} = \frac{1}{6}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The settings allow the special conditions in Eldar and Merhav's paper to hold, and $\gamma = \boldsymbol{\eta}^T(\mathbf{\Sigma}_{\max} - \mathbf{\Sigma}_{\min})\boldsymbol{\eta} = 2$. We would like to evaluate the maximum conditional risk associated with each

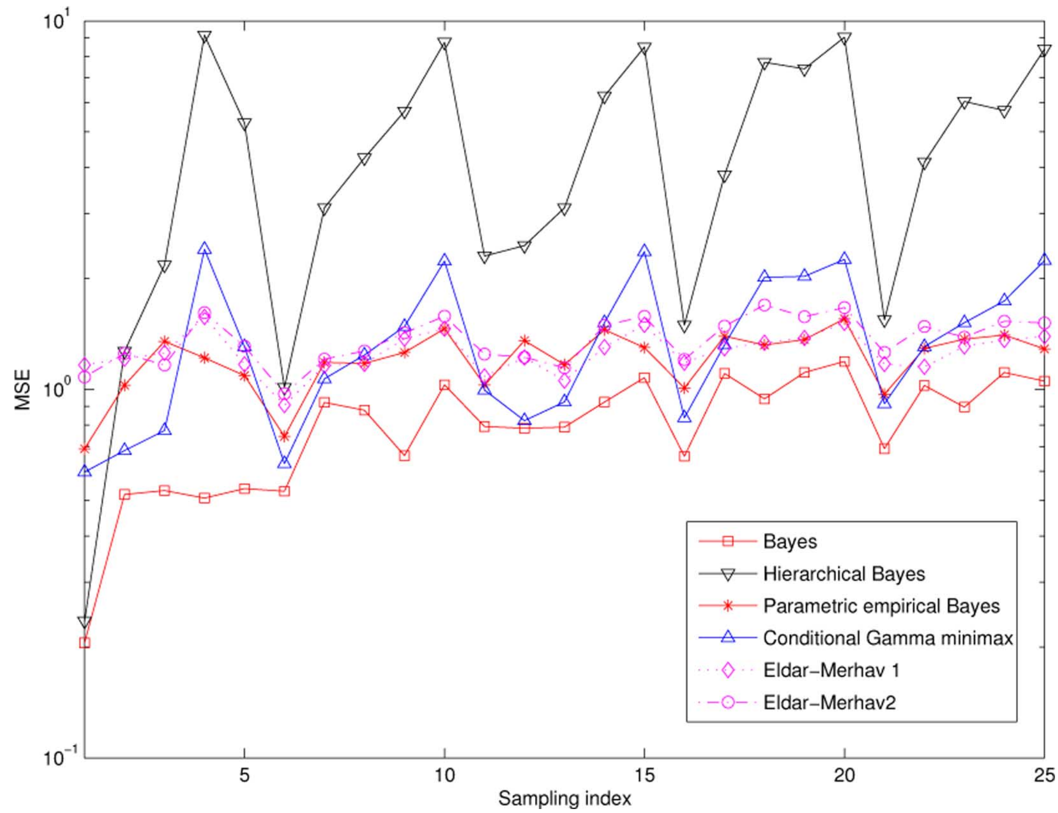

Fig. 7. MSE performance $K = 2$.

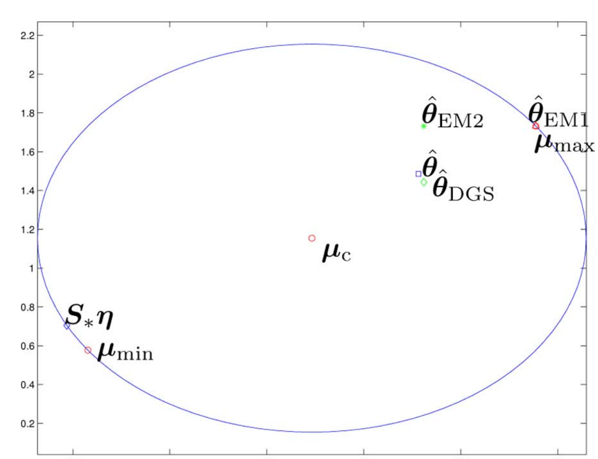

Fig. 8. Comparison between the true minimax, the DasGupta-Studden, and the Eldar and Merhav estimates. The maximum risks are, respectively, $g(\hat{\boldsymbol{\theta}}) = 8.426$, $g(\hat{\boldsymbol{\theta}}_{\mathrm{DGS}}) = 8.467$, $g(\hat{\boldsymbol{\theta}}_{\mathrm{EM}}) = 8.999$.

of the decisions on $\boldsymbol{\theta}$ by the above strategies as $\boldsymbol{\Sigma}$ varies between $\boldsymbol{\Sigma}_{\min}$ and $\boldsymbol{\Sigma}_{\max}$.

The estimates for this example are shown in Fig. 3. It is noted that $\hat{\boldsymbol{\theta}}_{\mathrm{DGS}}$ is quite close to $\hat{\boldsymbol{\theta}}$, whilst $\hat{\boldsymbol{\theta}}_{\mathrm{EM1}} = \boldsymbol{\mu}_{\max}$ incidentally. The associated maximum risks are shown in Table I. Clearly the strategies by Eldar and Merhav have higher maximum risk, which is not a surprise because their goal of optimality is different.

APPENDIX B
PROOF OF LEMMA 1

We note the same result was also stated in [9] without proof. For completeness, a proof is given here.

Proof: The conditional risk can be written as

$$\begin{aligned} E_{\boldsymbol{\theta}|\mathbf{y}}\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right] &= E_{\boldsymbol{\theta}|\mathbf{y}}\left[\hat{\boldsymbol{\theta}}^T\hat{\boldsymbol{\theta}}\right] - 2E_{\boldsymbol{\theta}|\mathbf{y}}\left[\hat{\boldsymbol{\theta}}^T\boldsymbol{\theta}\right] \\ &\quad + E_{\boldsymbol{\theta}|\mathbf{y}}\left[\boldsymbol{\theta}^T\boldsymbol{\theta}\right] \\ &= \hat{\boldsymbol{\theta}}^T\hat{\boldsymbol{\theta}} - 2\hat{\boldsymbol{\theta}}^T\boldsymbol{\mu} + E_{\boldsymbol{\theta}|\mathbf{y}}\left[\boldsymbol{\theta}^T\boldsymbol{\theta}\right] \end{aligned} \tag{25}$$

where it can be easily seen that the third term is

$$\begin{aligned} E_{\boldsymbol{\theta}|\mathbf{y}}\left[\boldsymbol{\theta}^T\boldsymbol{\theta}\right] &= E_{\boldsymbol{\theta}|\mathbf{y}}\left[\sum_{k=1}^K \theta_k^2\right] = \sum_{k=1}^K E_{\boldsymbol{\theta}|\mathbf{y}}\left[\theta_k^2\right] \\ &= \sum_{k=1}^K \left(\mu_k^2 + \Sigma_{kk}\right) \\ &= \boldsymbol{\mu}^T\boldsymbol{\mu} + \mathrm{tr}\left[\boldsymbol{\Sigma}\right]. \end{aligned} \tag{26}$$

The proof follows directly from (25) and (26) ■

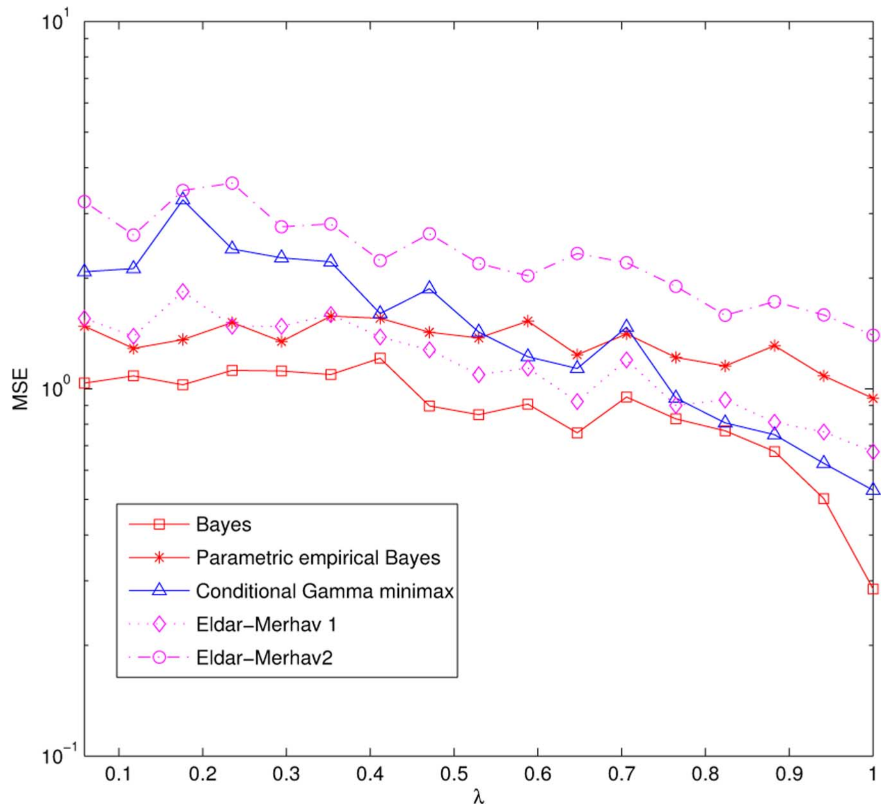

Fig. 9. Mean squared error (MSE) performance $K = 3$.

TABLE I
MAXIMUM CONDITIONAL RISK

Strategy	Maximum conditional risk
$\hat{\boldsymbol{\theta}}$	8.426
$\hat{\boldsymbol{\theta}}_{\rm DGS}$	8.467
$\hat{\boldsymbol{\theta}}_{\rm EM1}$	11.446
$\hat{\boldsymbol{\theta}}_{\rm EM2}$	8.999

APPENDIX C
PROOF OF LEMMA 2

Proof: First, we prove a somewhat more special case: If $\mathbf{z} = \boldsymbol{\Psi}\boldsymbol{\eta}$ with $\mathbf{0} \prec \boldsymbol{\Psi} \prec \mathbf{M}$ then $\mathbf{z}$ lies within the interior of the ellipsoid given by

$$\mathcal{E}_z : \left(\mathbf{z} - \frac{1}{2}\mathbf{M}\boldsymbol{\eta}\right)^T \mathbf{M}^{-1}\left(\mathbf{z} - \frac{1}{2}\mathbf{M}\boldsymbol{\eta}\right) \le \frac{1}{4}\boldsymbol{\eta}^T\mathbf{M}\boldsymbol{\eta}. \quad (27)$$

This can be directly obtained from Theorem 3 of Leamer [14] or by writing $\boldsymbol{\eta} = \boldsymbol{\Psi}^{-1}\mathbf{z}$ and deducing $\mathbf{z}^T\boldsymbol{\eta} > \mathbf{z}^T\boldsymbol{\Psi}^{-1}\mathbf{z}$, which leads to (27) with strict inequality after straightforward manipulations. For the converse of this special case, we pick an interior point $\mathbf{z}$ satisfying (27) (note strict $<$ is required) and it can be easily shown that

$$\mathbf{z}^T(\boldsymbol{\eta} - \mathbf{M}^{-1}\mathbf{z}) > 0.$$

Then according to Lemma 3 (see Appendix L) there exists a symmetric positive definite $\mathbf{D}$ such that

$$\mathbf{z} = \mathbf{D}(\boldsymbol{\eta} - \mathbf{M}^{-1}\mathbf{z})$$

from which

$$\mathbf{z} = (\mathbf{D}^{-1} + \mathbf{M}^{-1})^{-1}\boldsymbol{\eta}.$$

It can be easily seen that the choice $\boldsymbol{\Psi} = (\mathbf{D}^{-1} + \mathbf{M}^{-1})^{-1}$ satisfies $\mathbf{0} \prec \boldsymbol{\Psi} \prec \mathbf{M}$, which proves the converse.

Now we extend the result to the case $\mathbf{0} \preceq \boldsymbol{\Psi} \preceq \mathbf{M}$. Similar to the proof of Proposition 1, we observe that the set $\mathcal{S}_z$ of $\boldsymbol{\Psi}$ is a closed convex set, and we have established that the mapping of the interior of $\mathcal{S}_z$ is the interior of $\mathcal{E}_z$. As both these sets are convex and compact, and the mapping is linear, it follows from Lemma 5 that the mapping of $\mathcal{S}_z$ is the whole $\mathcal{E}_z$. The converse of this case follows from the converse of the above interior case and the continuity of the mapping. Note that this does not imply the boundary of $\mathcal{S}_z$ maps to the boundary of $\mathcal{E}_z$, but its closure does.

To apply this for our problem, we can rewrite

$$\boldsymbol{\mu} = (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta} + \boldsymbol{\Sigma}_{\min}\boldsymbol{\eta}$$

and apply the above result, including the converse, with $\mathbf{M} = \boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min}$ and $\mathbf{z} = \boldsymbol{\mu} - \boldsymbol{\Sigma}_{\min}\boldsymbol{\eta}$. ■

Remarks:

- In the above proof, we have only asserted that the mapping from $\mathcal{S}$ to the interior of $\mathcal{E}$ is *surjective*. For a point in the interior of $\mathcal{E}$, there may exist a family of ellipsoids passing through that point and bounded within $\mathcal{E}$. However, the statement in the above Lemma is sufficient for the result derived in the paper.
- Note that a trivial case is when $\boldsymbol{\mu}_0 = \mathbf{0}$ so that the ellipsoid $\mathcal{E}$ collapses to a single point $\mathbf{0}$. In this case, the problem is

TABLE II
NUMERICAL STUDY

$\hat{\boldsymbol{\theta}}$	$\boldsymbol{\Sigma}$	$\boldsymbol{\Sigma\eta}$	Boundary point?
$\begin{bmatrix} 1.4699 \\ 2.4054 \\ 1.0022 \end{bmatrix}$	$\begin{bmatrix} 2.3684 & -1.3147 & 0 \\ -1.3147 & 2.2632 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 0.57846 \\ 0.5222 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 1.3363 \\ 2.1381 \\ 0.93541 \end{bmatrix}$	$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1.6036 \\ 2.6726 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 1.6036 \\ 2.6726 \\ 1.069 \end{bmatrix}$	$\begin{bmatrix} 2.3684 & -1.3147 & 0 \\ -1.3147 & 2.2632 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 0.57846 \\ 0.5222 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 1.069 \\ 1.6036 \\ 1.3018 \end{bmatrix}$	$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1.6036 \\ 2.6726 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 2.069 \\ 1.6036 \\ 0.80178 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 0.53452 \\ 2.6726 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 1.069 \\ 2.6036 \\ 0.80178 \end{bmatrix}$	$\begin{bmatrix} 2.8947 & -0.63158 & 0 \\ -0.63158 & 1.2105 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1.2138 \\ 0.31352 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 1.069 \\ 3.6036 \\ 0.80178 \end{bmatrix}$	$\begin{bmatrix} 2.8947 & -0.63158 & 0 \\ -0.63158 & 1.2105 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1.2138 \\ 0.31352 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 1.069 \\ 1.6036 \\ 1.0518 \end{bmatrix}$	$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1.6036 \\ 2.6726 \\ 1.069 \end{bmatrix}$	YES
$\begin{bmatrix} 1.069 \\ 1.6036 \\ 1.3018 \end{bmatrix}$	$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1.6036 \\ 2.6726 \\ 1.069 \end{bmatrix}$	YES

trivial as $\boldsymbol{\mu} = \mathbf{0}$, independent of π and not considered in this work.

APPENDIX D
PROOFS OF THE RESULT (17)

The proof of (31) can be deduced in a similar manner to that of Lemma 2. First, consider the case $\mathbf{R}_{\boldsymbol{\theta\theta}} \succ \mathbf{0}$. It follows that $0 \prec \boldsymbol{\Sigma} \prec \boldsymbol{\Omega}^{-1}$. From $\boldsymbol{\mu} = \boldsymbol{\Sigma\eta}$, one can deduce $\boldsymbol{\mu}^T\boldsymbol{\eta} = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} > \boldsymbol{\mu}^T\boldsymbol{\Omega\mu}$. By substituting $\boldsymbol{\eta} = \boldsymbol{\Omega\mu}_0$, it follows that

$$\boldsymbol{\mu}^T\boldsymbol{\Omega\mu} - \boldsymbol{\mu}^T\boldsymbol{\Omega\mu}_0 < 0$$

from which one can easily arrive at (30) without the equality. To prove the equality part, we can use the same argument of convex sets and linear mapping as used in the proof of Lemma 2.

The ellipsoid for the case $\mathbf{R}_{\boldsymbol{\theta\theta}} \succeq \mathbf{R}_{\min}$ can be derived directly from Lemma 2 by noting that $\boldsymbol{\Sigma}_{\min} \preceq \boldsymbol{\Sigma} \preceq \boldsymbol{\Omega}^{-1}$. Hence, a direct application of Lemma 2 yields the following form:

$$(\mathcal{E}_l): \left(\boldsymbol{\mu} - \frac{1}{2}(\boldsymbol{\Omega}^{-1} + \boldsymbol{\Sigma}_{\min})\right)^T (\boldsymbol{\Omega}^{-1} - \boldsymbol{\Sigma}_{\min})^{-1} \times \left(\boldsymbol{\mu} - \frac{1}{2}(\boldsymbol{\Omega}^{-1} + \boldsymbol{\Sigma}_{\min})\right) \leq \frac{1}{4}\boldsymbol{\eta}^T(\boldsymbol{\Omega}^{-1} - \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta}. \quad (28)$$

Similarly, the ellipsoid for the case $\mathbf{R}_{\boldsymbol{\theta\theta}} \preceq \mathbf{R}_{\max}$ can be derived directly from Lemma 2 by noting that $\mathbf{0} \preceq \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_{\max}$. Hence, a direct application of Lemma 2 yields the following form:

$$(\mathcal{E}_u): \left(\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\Sigma}_{\max}\boldsymbol{\eta}\right)^T \boldsymbol{\Sigma}_{\max}^{-1} \left(\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\Sigma}_{\max}\boldsymbol{\eta}\right) \leq \frac{1}{4}\boldsymbol{\eta}^T\boldsymbol{\Sigma}_{\max}\boldsymbol{\eta}. \quad (29)$$

We note that in all cases, any ellipsoid can be written as an inequality $f(\boldsymbol{\mu}) \leq 0$, where $f(\boldsymbol{\mu})$ is a quadratic function of $\boldsymbol{\mu}$. If $f(\boldsymbol{\mu}) = 0$ then $\boldsymbol{\mu}$ is on the surface of the ellipsoid; otherwise, it is an interior point. Denote such quadratic functions for $\mathcal{E}$, $\mathcal{E}_u$, $\mathcal{E}_l$ as $f_{\mathcal{E}}$, $f_{\mathcal{E}_u}$, $f_{\mathcal{E}_l}$. We prove (17) in two steps.

- Prove that $f_{\mathcal{E}}(\boldsymbol{\mu}_{\min}) = f_{\mathcal{E}_l}(\boldsymbol{\mu}_{\min}) = \mathbf{0}$ and $f_{\mathcal{E}}(\boldsymbol{\mu}_{\max}) = f_{\mathcal{E}_u}(\boldsymbol{\mu}_{\max}) = \mathbf{0}$, i.e., $V_{\min}$ and $V_{\max}$ are the surface points where the ellipsoids intersect.
- Prove that the normal vectors of the tangent planes for $\mathcal{E}$ and $\mathcal{E}_l$ at $V_{\min}$ are the same. Similarly, the normal vectors of the tangent planes for $\mathcal{E}$ and $\mathcal{E}_u$ at $V_{\max}$ are the same.

The first step is a simple substitution. For the second step, we recall a basic result that the normal of a tangent plane for a surface $f(\boldsymbol{\mu}) = 0$ at $\boldsymbol{\mu}$ is $C\nabla f(\boldsymbol{\mu})$ where C is a normalization constant. As it can be easily verified that

$$\nabla f_{\mathcal{E}}(\boldsymbol{\mu}_{\min}) = \nabla f_{\mathcal{E}_l}(\boldsymbol{\mu}_{\min}) = \boldsymbol{\eta}$$
$$\nabla f_{\mathcal{E}}(\boldsymbol{\mu}_{\max}) = \nabla f_{\mathcal{E}_u}(\boldsymbol{\mu}_{\max}) = \boldsymbol{\eta}$$

the proof follows.

APPENDIX E
PROOF OF PROPOSITION 3

Proof: From Lemma 2, for any point $\boldsymbol{\mu} \in \mathrm{bd}(\mathcal{E})$, it is always possible to find some $\boldsymbol{\Sigma} \in \mathrm{bd}(\mathcal{S})$ such that $\boldsymbol{\mu} = \boldsymbol{\Sigma\eta}$. Suppose that $\boldsymbol{\Sigma} \notin \mathcal{S}_*$, then it is either in $\mathcal{S}_{\min} \setminus \mathcal{S}_*$ or $\mathcal{S}_{\max} \setminus \mathcal{S}_*$ (i.e., on the boundary of $\mathcal{S}$; otherwise, it would map to an interior point of $\mathcal{E}$ according to Lemma 2). Suppose that $\boldsymbol{\Sigma} \in \{\mathcal{S}_{\min} \setminus \mathcal{S}_*\}$. The existence of a point in $\mathcal{S}_*$ is based on the following facts.

- It is possible to find some $\lambda > 1$ such that

$$\mathbf{\Sigma}_1 = \mathbf{\Sigma}_{\min} + \lambda(\mathbf{\Sigma} - \mathbf{\Sigma}_{\min}) \tag{30}$$

  belongs to $\mathcal{S}_*$ (i.e., extending the line connecting $\mathbf{\Sigma}_{\min}$ and $\mathbf{\Sigma}$ on the surface of the cone $\mathcal{S}_{\min}$ should intercept $\mathcal{S}_*$ at some $\mathbf{\Sigma}_1$)
- We note that it is possible to move all the terms in the equation of ellipsoid $\mathcal{E}$ (cf. (12)) to the left-hand side and obtain a quadratic function in terms of $\mathbf{\Sigma}$. This ellipsoid function is convex and the maximum is 0 (found at the boundary where the equality occurs).
- $\mathbf{\Sigma}$ lies between $\mathbf{\Sigma}_{\min}$ and $\mathbf{\Sigma}_1$, and the values of the ellipsoid function at $\mathbf{\Sigma}_{\min}$ and $\mathbf{\Sigma}$ are zero.

From this result, we can easily deduce that the ellipsoid function evaluated at $\mathbf{\Sigma}_1$ is also zero[4], which implies $\mathbf{\Sigma}_1$ also maps to a boundary point of $\mathcal{E}$. A similar argument can be made for the case when $\mathbf{\Sigma} \in \{\mathcal{S}_{\max} \setminus \mathcal{S}_*\}$. It remains to prove the case when $\boldsymbol{\mu}$ is at either $\boldsymbol{\mu}_{\min}$ or $\boldsymbol{\mu}_{\max}$. The proof for this case follows from the fact that the set $\mathcal{S}_*$ is compact and the mapping from $\mathcal{S}$ to $\mathcal{E}$ is linear. In other words, we can consider, for example, a sequence $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \ldots$ on the boundary of $\mathcal{E}$ that converges to $\boldsymbol{\mu}_{\min}$. Then, as $\mathcal{S}_*$ is compact, and due to the above result, there exists a corresponding convergent subsequent $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \ldots$ in $\mathcal{S}_*$ with a limit point $\mathbf{\Sigma}_* \in \mathcal{S}_*$. It follows that $\mathbf{\Sigma}_*$ maps to $\boldsymbol{\mu}_{\min}$. Similar argument can be made for $\boldsymbol{\mu}_{\max}$. This establishes that there exists $\mathcal{S}_*^b \subseteq \mathcal{S}_*$ that maps to $\mathsf{bd}(\mathcal{E})$. To prove the uniqueness of the mapping, suppose that there exist two distinct $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2 \in \mathcal{S}_*^b$ that maps to the same boundary point of $\mathcal{E}$. It follows that all convex combinations $\lambda\mathbf{\Sigma}_1 + (1-\lambda)\mathbf{\Sigma}_2, \lambda \in [0, 1]$ map to that boundary point. However, there exists a convex combination that yields the interior point of $\mathcal{S}$, meaning this interior point maps to a boundary point of $\mathcal{E}$, which contradicts Lemma 2.

The result above does not rule out the case that some $\mathbf{\Sigma} \in \mathcal{S}_*$ could map to an interior point of $\mathcal{E}$. However, for $K = 2$, we show that the mapping from $\mathcal{S}_*$ is exclusively to $\mathsf{bd}(\mathcal{E})$. Similar to the proof of Lemma 2, it suffices to consider the special case where $\mathbf{\Psi} \succsim \mathbf{0}$ and $\mathbf{\Psi} \precsim \mathbf{M}$ (this corresponds to the set $\mathcal{S}_*$ in the generalized case). Substituting $\mathbf{z} = \mathbf{\Psi}\boldsymbol{\eta}$ into the equation of ellipsoid $\mathcal{E}_z$ we need to show that

$$\mathbf{z}^T(\mathbf{M}^{-1}\mathbf{z} - \boldsymbol{\eta}) = 0$$

or equivalently

$$\boldsymbol{\eta}^T \left(\mathbf{\Psi} - \mathbf{\Psi}\mathbf{M}^{-1}\mathbf{\Psi}\right) \boldsymbol{\eta} = 0.$$

Due to Corollary 2 (see Appendix L), the middle term is $\mathbf{0}$ which concludes the proof.

We also note that for the case $K \geq 3$ the result of Corollary 3 specifies the subset $\mathcal{S}_*^b$. ■

[4]Suppose that the quadratic function is h which is convex and ≤ 0 and such that $h(\mathbf{\Sigma}_1) < 0$. Then as $h(\mathbf{\Sigma}_{\min}) = 0$, its convexity implies that $h(\mathbf{\Sigma}) < 0$ which is a contradiction. Hence, $h(\mathbf{\Sigma}_1) = 0$

APPENDIX F
PROOF OF COROLLARY 1

Proof: The first result follows from the fact that $\mathsf{tr}[\mathbf{\Sigma}] \geq \mathsf{tr}[\mathbf{\Sigma}_{\min}], \forall \mathbf{\Sigma} \in \mathcal{S}_*$[5]. Also, it is always possible to find at least a point on the boundary of $\mathcal{E}$ that has the same distance to any $\hat{\boldsymbol{\theta}}$ as that of $\boldsymbol{\mu}_{\min}$ (the trivial point is $V_{\min}$ itself; note that from Proposition 3 there exists $\mathbf{\Sigma}_s \in \mathcal{S}_*$ that maps to $\boldsymbol{\mu}_{\min}$). Hence

$$g_*(\hat{\boldsymbol{\theta}}) \geq \underbrace{\| \hat{\boldsymbol{\theta}} - \Sigma_s \boldsymbol{\eta} \|^2}_{= \|\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\min}\|^2} + \underbrace{\mathsf{tr}[\mathbf{\Sigma}]}_{\geq \mathsf{tr}[\mathbf{\Sigma}_{\min}]} \geq g_{\min}(\hat{\boldsymbol{\theta}}).$$

For the second result, apart from directly applying the first result we need to prove that, in this case, the minimax solution cannot be found outside the ellipsoid $\mathcal{E}$. From (21), we note that each $f(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}_i)$ is a convex and quadratic function of $\hat{\boldsymbol{\theta}}$ and attains its minimum at a point on the boundary of $\mathcal{E}$. It follows that the minimum of the convex function $g_*(\hat{\boldsymbol{\theta}})$ must be found within $\mathcal{E}$. Finally, note that $g_{\max}(\hat{\boldsymbol{\theta}})$ is also a quadratic and convex function with a minimum at $\boldsymbol{\mu}_{\max}$ (Intuitively, $\mathcal{E}$ is the space of the posterior mean, so the minimax solution must be found here). ■

APPENDIX G
PROOF OF THEOREM 1

From (9), we note that the supremum problem does not yield a conditional risk less than $g_{\max}(\boldsymbol{\mu}_{\max}) = \mathsf{tr}[\mathbf{\Sigma}_{\max}]$. Hence, the problem is that of showing that the supremum over $\mathbf{\Sigma}$ for $f(\boldsymbol{\mu}_{\max}, \mathbf{\Sigma})$ is actually $\mathsf{tr}[\mathbf{\Sigma}_{\max}]$ given $\boldsymbol{\eta}^T(\mathbf{\Delta})\boldsymbol{\eta} \leq 1$, where $\mathbf{\Delta} = \mathbf{\Sigma}_{\max} - \mathbf{\Sigma}$ with $\mathbf{\Sigma}_{\max} \succeq \mathbf{\Delta} \succeq \mathbf{0}$. This is equivalent to showing $\boldsymbol{\eta}^T\mathbf{\Delta}^2\boldsymbol{\eta} \geq \mathsf{tr}[\mathbf{\Delta}]$ with $\boldsymbol{\eta}^T\mathbf{\Delta}\boldsymbol{\eta} \leq 1$. As $\mathbf{\Delta}$ is symmetric and positive semi-definite, the eigenvalue decomposition yields $\mathbf{\Delta} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ with $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ where $\lambda_i \geq 0$. Using the property that $\mathsf{tr}[\mathbf{\Delta}] = \sum \lambda_i$, and defining $\mathbf{v} = \mathbf{U}^T\boldsymbol{\eta}$, the problem is equivalent to showing $\sum_i \lambda_i(\lambda_i v_i^2 - 1) \leq 0$ given $\sum_i \lambda_i v_i^2 \leq 1$. The required results follows because each of the term in the sum ≤ 0.

APPENDIX H
PROOF OF PROPOSITION 4

The first part of the result follows as $g_{\max}(\boldsymbol{\mu}_{\min}) \geq g_*(\boldsymbol{\mu}_{\min})$, $g_{\max}(\boldsymbol{\mu}_{\max}) < g_*(\boldsymbol{\mu}_{\max})$, and $g(\hat{\boldsymbol{\theta}})$ is both convex and continuous. The second part of the result is based on the property of the minimax solution and can be easily proved by contradiction. To see this, we start from the first part, which proves the existence of $\mathcal{E}_1$ and $\mathcal{E}_2$. From the proof of the existence, we know that $\boldsymbol{\mu}_{\max} \in \mathcal{E}_1$. Suppose, on the contrary, that the minimax solution, denoted by $\hat{\boldsymbol{\theta}}_*$, does not lie in $\mathcal{E}_1 \cap \mathcal{E}_2$. There are two possible scenarios.

- $\hat{\boldsymbol{\theta}}_* \in \mathcal{E}_2 \setminus \mathcal{E}_1 \cap \mathcal{E}_2$. Then we know from the first part that $g_{\max}(\boldsymbol{\theta}) \geq g_*(\hat{\boldsymbol{\theta}}) \forall \hat{\boldsymbol{\theta}} \in \mathcal{E}_2 \setminus \mathcal{E}_1 \cap \mathcal{E}_2$. However, the convexity implies $g_{\max}(\boldsymbol{\theta}) \geq g_{\max}(\boldsymbol{\mu}_{\max})$ which means that $\hat{\boldsymbol{\theta}}_* = \boldsymbol{\mu}_{\max}$. This is the case when $\gamma \leq 1$, which is a contradiction.
- $\hat{\boldsymbol{\theta}}_* \in \mathcal{E}_1 \setminus \mathcal{E}_1 \cap \mathcal{E}_2$. Then $\hat{\boldsymbol{\theta}}_*$ must be a local minimum of $g_*(\hat{\boldsymbol{\theta}})$ in $\mathcal{E}_1$. The convexity of $g_*(\hat{\boldsymbol{\theta}})$ implies that $\forall \hat{\boldsymbol{\theta}} \in \mathcal{E}_1 \cap$

[5]Note: A stronger result is strictly $>$ for regular cases.

$\mathcal{E}_2$ and hence $g_*(\hat{\boldsymbol{\theta}}) \geq g_*(\hat{\boldsymbol{\theta}}_*)$. Specifically, we can always find $\hat{\boldsymbol{\theta}} \in \mathcal{E}_1 \cap \mathcal{E}_2$ such that $\hat{\boldsymbol{\theta}}$ is a convex combination of $\hat{\boldsymbol{\theta}}_*$ and $\boldsymbol{\mu}_{\max}$. It follows from the convexity of $g_{\max}(\hat{\boldsymbol{\theta}})$ that $g_{\max}(\hat{\boldsymbol{\theta}}) \leq g_{\max}(\hat{\boldsymbol{\theta}}_*)$. However, we have $g_{\max}(\hat{\boldsymbol{\theta}}) = g_*(\hat{\boldsymbol{\theta}})$. This leads to a contradiction.

The two scenarios both lead to a contradiction, which concludes the proof.

Appendix I
Proof of Proposition 5

We note that

$$\frac{\partial f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\theta}} = 2(\boldsymbol{\theta} - \boldsymbol{\Sigma}\boldsymbol{\eta}).$$

Consider a small neighborhood $\boldsymbol{\theta}' = \hat{\boldsymbol{\theta}} + \delta \mathbf{z}$ where $\mathbf{z}$ is some unit vector and δ is a small number. Using first-order approximation, one obtains

$$f(\boldsymbol{\theta}', \boldsymbol{\Sigma}_{\max}) = f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\max}) + 2\delta(\hat{\boldsymbol{\theta}} - \boldsymbol{\Sigma}_{\max}\boldsymbol{\eta})^T \mathbf{z}$$
$$f(\boldsymbol{\theta}', \boldsymbol{\Sigma}_*) = f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_*) + 2\delta(\hat{\boldsymbol{\theta}} - \boldsymbol{\Sigma}_*\boldsymbol{\eta})^T \mathbf{z}.$$

As $\hat{\boldsymbol{\theta}}$ is a convex combination of $\boldsymbol{\mu}_{\max}$ and $\boldsymbol{\Sigma}\boldsymbol{\eta}$, it follows that: $(\hat{\boldsymbol{\theta}} - \boldsymbol{\Sigma}_{\max}\boldsymbol{\eta})^T(\hat{\boldsymbol{\theta}} - \mathit{\Sigma}_*\boldsymbol{\eta}) \geq 0$. Thus

$$\max\{f(\boldsymbol{\theta}', \boldsymbol{\Sigma}_{\max}), f(\boldsymbol{\theta}', \boldsymbol{\Sigma}_*)\} \geq g(\hat{\boldsymbol{\theta}}).$$

As $g\left(\boldsymbol{\theta}'\right) \geq \max\left\{f(\boldsymbol{\theta}', \boldsymbol{\Sigma}_{\max}), f(\boldsymbol{\theta}', \boldsymbol{\Sigma}_*)\right\}$, it follows that $g(\boldsymbol{\theta}') \geq g(\hat{\boldsymbol{\theta}})$ in a small neighborhood of $\boldsymbol{\theta}$. As $g(\boldsymbol{\theta})$ is convex, this implies $\hat{\boldsymbol{\theta}}$ is the minimax solution.

Appendix J
Proof of Theorem 3

For notational brevity, we introduce $\mathbf{D} = \boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min}$. We rewrite

$$\boldsymbol{\Sigma} = \frac{1}{2}\left(\boldsymbol{\Sigma}_{\max} + \boldsymbol{\Sigma}_{\min}\right) + \frac{1}{2}\mathbf{E} \tag{31}$$

where $\mathbf{E}$ is some symmetric positive semi-definite matrix. The condition $\boldsymbol{\Sigma}_{\min} \preceq \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_{\max}$ implies

$$-\mathbf{D} \preceq \mathbf{E} \preceq \mathbf{D}. \tag{32}$$

Using the same approach as in [9], we show that the choice (20) satisfies

$$f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\max}) = f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_*) = \sup_{\boldsymbol{\Sigma} \in \mathcal{S}} f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma}_* = (\boldsymbol{\Sigma}_{\max} + \boldsymbol{\Sigma}_{\min})/2 + \mathbf{E}_*/2$ and

$$\mathbf{E}_* = \mathbf{D} - \frac{2\mathbf{D}\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{D}}{\boldsymbol{\eta}^T\mathbf{D}\boldsymbol{\eta}}.$$

First, we show that

$$\sup_{\boldsymbol{\Sigma} \in \mathcal{S}_*} f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) = f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\max}).$$

By substituting (31) into the formula for $f(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$, as shown in (9), this is equivalent to showing that

$$\frac{1}{4}\boldsymbol{\eta}^T\mathbf{E}^2\boldsymbol{\eta} - \frac{1}{2}\frac{\boldsymbol{\eta}^T\mathbf{D}\mathbf{E}\boldsymbol{\eta}}{\boldsymbol{\eta}^T\mathbf{D}\boldsymbol{\eta}} + \frac{1}{2}\mathrm{tr}[\mathbf{E}] \leq \frac{1}{4}\boldsymbol{\eta}^T\mathbf{D}^2\boldsymbol{\eta} - \frac{1}{2}\frac{\boldsymbol{\eta}^T\mathbf{D}^2\boldsymbol{\eta}}{\boldsymbol{\eta}^T\mathbf{D}\boldsymbol{\eta}} + \frac{1}{2}\mathrm{tr}[\mathbf{D}]. \tag{33}$$

We prove (33) using the following arguments.

- $\frac{1}{4}\boldsymbol{\eta}^T\mathbf{E}^2\boldsymbol{\eta} + \leq \frac{1}{4}\boldsymbol{\eta}^T\mathbf{D}^2\boldsymbol{\eta}$ is a direct consequence of (32) as $\boldsymbol{\eta}^T\mathbf{E}^2\boldsymbol{\eta} \leq \lambda^2_{\max}(\mathbf{E})\|\boldsymbol{\eta}\|_2^2 \leq \lambda^2_{\max}(\mathbf{D})\|\boldsymbol{\eta}\|_2^2 = \boldsymbol{\eta}^T\mathbf{D}^2\boldsymbol{\eta}$, where the last equality follows from the assumption that $\boldsymbol{\eta}$ lies in the subspace of the top eigenvectors whose eigenvalues are equal.
- Using Lemma 6, the trick $\boldsymbol{\eta}^T\mathbf{D}\boldsymbol{\eta} = \mathrm{tr}[\mathbf{D}\boldsymbol{\eta}\boldsymbol{\eta}^T]$, and the trace property $\mathrm{tr}[\mathbf{AB}] = \mathrm{tr}[\mathbf{BA}]$, it can be easily shown that

  $$\frac{\boldsymbol{\eta}^T\mathbf{D}(\mathbf{D}-\mathbf{E})\boldsymbol{\eta}}{\boldsymbol{\eta}^T\mathbf{D}\boldsymbol{\eta}} \leq \mathrm{tr}[\mathbf{D}-\mathbf{E}]$$

  which implies, after straightforward manipulations, that

  $$\mathrm{tr}[\mathbf{E}] - \frac{\boldsymbol{\eta}^T\mathbf{D}\mathbf{E}\boldsymbol{\eta}}{\boldsymbol{\eta}^T\mathbf{D}\boldsymbol{\eta}} \leq \mathrm{tr}[\mathbf{D}] - \frac{\boldsymbol{\eta}^T\mathbf{D}^2\boldsymbol{\eta}}{\boldsymbol{\eta}^T\mathbf{D}\boldsymbol{\eta}}.$$

The equalities occur at a trivial choice $\mathbf{E} = \mathbf{D}$. However, it is also possible to show, after tedious verification, that the equalities also occur at $\mathbf{E} = \mathbf{E}_*$. We omit the detail.

It is also of interest to note that $\boldsymbol{\Sigma}_*\boldsymbol{\eta} = \boldsymbol{\mu}_{\min}$. Thus, it is easy to verify that the choice for $\hat{\boldsymbol{\theta}}$ in (20) is a convex combination of $\boldsymbol{\mu}_{\max} = \boldsymbol{\Sigma}_{\max}\boldsymbol{\eta}$ and $\boldsymbol{\mu}_{\min} = \boldsymbol{\Sigma}_*\boldsymbol{\eta}$. It follows by Proposition 5 that (20) is the minimax solution.

Appendix K
SDP Formulation

With the new variable

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{bmatrix}$$

where $\mathbf{X}_{11} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\min}$, $\mathbf{X}_{22} = \boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}$, and $\mathbf{X}_{21} = \mathbf{X}_{12} = \mathbf{0}$, the problem (22) is equivalent to

$$\begin{cases} \min & \mathrm{tr}[\mathbf{CX}] \\ \text{s.t.} & \mathbf{X} \succeq \mathbf{0}, \\ & \mathbf{X}_{11}\boldsymbol{\eta} = \boldsymbol{\mu} - \boldsymbol{\mu}_{\min} \\ & \mathbf{X}_{12} = \mathbf{X}_{21} = \mathbf{0} \\ & \mathbf{X}_{11} + \mathbf{X}_{22} = \boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min} \end{cases}$$

where $\mathbf{C} = \begin{bmatrix} -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$. This is exactly the form of a SDP problem and thus can be considered solved.

In practice, many SDP solvers, for example SeDuMi and Yalmip [22], [16], can handle multiple positive semi-definite constraints already and thus (22) can be easily implemented by a few lines of code.

APPENDIX L
SOME SUPPLEMENTARY RESULTS

Lemma 3: Assume that $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\mu}^T\mathbf{x} > 0$. Then there exists a symmetric positive definite matrix $\mathbf{D}$ which satisfies

$$\boldsymbol{\mu} = \mathbf{D}\mathbf{x}.$$

Proof: It can be easily seen that this equation is true for $n = 1$. We therefore need to prove the result for the case $n \geq 2$. The idea is to project a solution on $\mathbb{R}^2$ to $\mathbb{R}^n$. Let

$$\mathbf{z} = \frac{\mathbf{x}}{\|\mathbf{x}\|}, \ \mathbf{v} = \boldsymbol{\mu} - \mathbf{z}^T\boldsymbol{\mu}\mathbf{z}.$$

It can be easily seen that $\mathbf{v} \perp \mathbf{x}$ (i.e., $\mathbf{v}^T\mathbf{x} = 0$) and that $\mathsf{span}(\boldsymbol{\mu}, \mathbf{v}) = \mathsf{span}(\boldsymbol{\mu}, \mathbf{x})$. We note that $\boldsymbol{\mu}^T\mathbf{x} > 0$. Therefore, there always exists a symmetric positive semi-definite matrix

$$\mathbf{D}_2 = \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

with $a = 1/(\boldsymbol{\mu}^T\mathbf{x}),\ c > 0,\ b^2 < ac$, such that

$$\mathbf{D}_2 \begin{bmatrix} \boldsymbol{\mu}^T\mathbf{x} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Let $\mathbf{U} = [\mathbf{u_1}, \ldots, \mathbf{u}_{n-2}]$ be $n-2$ linearly independent vectors in the orthogonal subspace of $[\boldsymbol{\mu} \quad \mathbf{v}]$

$$\mathsf{span}(\mathbf{u_1}, \ldots, \mathbf{u}_{n-2}) = \{\mathbb{R}^n \setminus \mathsf{span}(\boldsymbol{\mu}, \mathbf{v})\}.$$

Denote $\mathbf{W} = [\boldsymbol{\mu} \quad \mathbf{v}]$. It is easy to show that the matrix

$$\mathbf{D} = [\mathbf{W} \quad \mathbf{U}] \begin{bmatrix} \mathbf{D}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-2} \end{bmatrix} \begin{bmatrix} \mathbf{W}^T \\ \mathbf{U}^T \end{bmatrix}.$$

is symmetric positive definite. Further

$$\begin{aligned} \mathbf{D}\mathbf{x} &= \mathbf{W}\mathbf{D}_2\mathbf{W}^T\mathbf{x} = [\boldsymbol{\mu}\, \mathbf{v}]\mathbf{D}_2 \begin{bmatrix} \boldsymbol{\mu}^T\mathbf{x} \\ \mathbf{v}^T\mathbf{x} \end{bmatrix} \\ &= [\boldsymbol{\mu}\, \mathbf{v}] \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \boldsymbol{\mu} \end{aligned} \tag{34}$$

where we have used $\mathbf{U}^T\mathbf{x} = \mathbf{0}$ in (34). This concludes the proof. ∎

Lemma 4: Consider two symmetric matrices $\mathbf{A} \succ \mathbf{0}, \mathbf{B} \succeq \mathbf{0}$ with size $K \times K$. Suppose that $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$. We have

$$\mathbf{X} = \mathbf{B} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B} \succeq \mathbf{0}.$$

Proof: From the given condition, it follows that there exists a transformation $\mathbf{W}$ that simultaneously diagonalizes $\mathbf{A}$ and $\mathbf{B}$ (see [11, p. 469] for an algorithm)

$$\mathbf{W}^T\mathbf{A}\mathbf{W} = \mathbf{I},\ \mathbf{W}^T\mathbf{B}\mathbf{W} = \mathbf{C},$$

where $\mathbf{C} = \mathsf{diag}(c_1, \ldots, c_K)$, $c_k \geq 0$, $k = 1, \ldots, K$, is a diagonal matrix. It follows that $\mathbf{I} \succeq \mathbf{C}$ which results in $0 \leq c_k \leq 1$, $k = 1, \ldots, K$. Proving the Lemma amounts to proving that $\mathbf{W}^T\mathbf{X}\mathbf{W} \succeq \mathbf{0}$, which is easily translated to $\mathbf{W}^T(\mathbf{C} - \mathbf{C}^2)\mathbf{W} \succeq \mathbf{0}$. This result can be easily verified. ∎

Corollary 2: Given the conditions in Lemma 4, and that $\mathbf{B} \succsim \mathbf{0}$, $\mathbf{A} \succsim \mathbf{B}$ and $K = 2$, it follows that $\mathbf{X} = \mathbf{0}$.

Proof: In this case, $\mathbf{C} = \mathsf{diag}(c_1, c_2)$. Combining the two conditions $\mathbf{B} \succsim \mathbf{0}$ and $\mathbf{A} \succsim \mathbf{B}$, we can deduce that one diagonal element is 0, while the other is 1. Verification that $\mathbf{C} - \mathbf{C}^2 = \mathbf{0}$ is straightforward. ∎

Corollary 3: Given the conditions in Lemma 4. It follows that $\mathbf{X} = \mathbf{0}$ if a nontrivial $\mathbf{B}$ can be decomposed as

$$\mathbf{B} = \mathbf{W}^{-T}\mathbf{C}_0\mathbf{W}^{-1}, \ \text{with} \ \mathbf{W} = \boldsymbol{\Lambda}_{\mathrm{A}}^{-1/2}\mathbf{U}_{\mathrm{A}}^T\mathbf{U}_k$$

where $\boldsymbol{\Lambda}_{\mathrm{A}}$ and $\mathbf{U}_{\mathrm{A}}$ are the eigenvalue and eigenvector matrices of $\mathbf{A}$, $\mathbf{C}_0$ is some diagonal matrix with 1's and 0's in the diagonal, and $\mathbf{U}_k$ is a unitary matrix.

Proof: It is easy to verify that $\mathbf{W}$ simultaneously diagonalizes $\mathbf{A}$ and $\mathbf{B}$, as in the proof of Lemma 4, and that $\mathbf{C}_0 - \mathbf{C}_0^2 = \mathbf{0}$. ∎

Lemma 5: Let X and Y be two finite-dimensional linear spaces. If there exist two sets $\mathcal{S}$ in X, bounded, and $\mathcal{E}$ in Y, and a linear mapping $f : X \mapsto Y$ such that

$$f(\mathcal{S}) = \mathcal{E} \tag{35}$$

then $f(\mathsf{cl}(\mathcal{S})) = \mathsf{cl}(\mathcal{E})$, where cl denotes the closure of a set.

Proof: From the given information, it follows that f is continuous, hence $f(\mathsf{cl}(\mathcal{S})) \subseteq \mathsf{cl}(\mathcal{E})$. Further, $\mathcal{S}$ is bounded, therefore $\mathsf{cl}(\mathcal{S})$ is bounded and thus compact. So from (35), we have

$$\mathcal{E} \subseteq f(cl(\mathcal{S})) \subseteq cl(\mathcal{E}).$$

On the other hand, a continuous function maps a compact set to a compact set, so $f(\mathsf{cl}(\mathcal{S}))$ must be compact and therefore closed. But, by definition, $\mathsf{cl}(\mathcal{E})$ is the smallest closed set that contains $\mathcal{E}$, so $\mathsf{cl}(\mathcal{E}) \subseteq f(\mathsf{cl}(\mathcal{S}))$. All these imply $\mathsf{cl}(\mathcal{E}) = f(\mathsf{cl}(\mathcal{S}))$. ∎

Lemma 6: For two positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$ of compatible dimensions, it is the case that

$$\mathsf{tr}[\mathbf{A}]\mathsf{tr}[\mathbf{B}] \geq \mathsf{tr}[\mathbf{A}\mathbf{B}].$$

Proof: This result is a direct consequence of classical results on eigenvalues of positive semi-definite matrices [13]

$$\mathsf{tr}[\mathbf{A}\mathbf{B}] \leq \sum_i \lambda_i(\mathbf{A})\lambda_i(\mathbf{B}).$$

As both $\mathbf{A}$ and $\mathbf{B}$ are positive semi-definite, their eigenvalues are nonnegative. Thus

$$\mathsf{tr}[\mathbf{A}\mathbf{B}] \leq \sum_i \lambda_i(\mathbf{A}) \sum_i \lambda_i(\mathbf{B}) = \mathsf{tr}[\mathbf{A}]\mathsf{tr}[\mathbf{B}].$$

One trivial case, which will be useful in the main paper, is that the inequality occurs when $\mathbf{A}$ is full-rank and $\mathrm{rank}(\mathbf{B}) = 1$. ∎

APPENDIX M
NUMERICAL EXAMPLE

In the paper, we give a conjecture that if $\boldsymbol{\Sigma}$ is the maximizer of

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\Sigma}\boldsymbol{\eta}\|_2^2 + \mathsf{tr}[\boldsymbol{\Sigma}]$$

then $\boldsymbol{\Sigma}\boldsymbol{\eta}$ is a boundary point of the ellipsoid $\mathcal{E}$. In what follows, we present a numerical study that supports this conjecture. We consider the case $K = 3$. As the above objective is a convex function of $\boldsymbol{\Sigma}$, we need to use brute-force search over the set $\mathcal{S} = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma}_{\min} \preceq \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_{\max}\}$. To verify if $\boldsymbol{\Sigma}$ maps to a boundary point, we need to check if

$$(\boldsymbol{\Sigma}\boldsymbol{\eta} - \boldsymbol{\mu}_c)^T(\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min})^{-1}(\boldsymbol{\Sigma}\boldsymbol{\eta} - \boldsymbol{\mu}_c) - \frac{1}{4}\boldsymbol{\eta}^T(\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta} = 0$$

where $\boldsymbol{\mu}_c = (\boldsymbol{\Sigma}_{\max} + \Sigma_{\min})\boldsymbol{\eta}/2$.

In this numerical study, we set

$$\boldsymbol{\Sigma}_{\min} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\max} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \boldsymbol{\eta} = \sqrt{\frac{2}{7}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

This choice ensures that $\gamma = \boldsymbol{\eta}^T(\boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min})\boldsymbol{\eta} = 2 > 1$ and that $\boldsymbol{\eta}$ does not lie in the principal subspace of $\mathbf{D} = \boldsymbol{\Sigma}_{\max} - \boldsymbol{\Sigma}_{\min}$. In other words, the choice is not special.

The special points of the ellipsoid $\mathcal{E}$ are

$$\boldsymbol{\mu}_{\min} = \begin{bmatrix} 0.5345 \\ 0.5345 \\ 0.5345 \end{bmatrix}, \quad \boldsymbol{\mu}_{\max} = \begin{bmatrix} 1.6036 \\ 2.6726 \\ 1.0690 \end{bmatrix}, \quad \boldsymbol{\mu}_c = \begin{bmatrix} 1.0690 \\ 1.6036 \\ 0.0818 \end{bmatrix}.$$

Table II shows the brute-force search results for several values of $\hat{\boldsymbol{\theta}}$ uniformly selected within the ellipsoid $\mathcal{E}$. Clearly, with all the points $\hat{\boldsymbol{\theta}}$ that we picked, the maximizer $\boldsymbol{\Sigma}$ is either at $\boldsymbol{\Sigma}_{\max}$ or maps to another boundary point.

Further, to support the conjecture for the case $K = 3$, we repeat the numerical study of MSE with the following parameters:

$$\mathbf{C} = \begin{bmatrix} 3 & 0.5 & 0.2 \\ -0.1 & 1 & -0.3 \\ 0.1 & 0.2 & 1.5 \end{bmatrix}, \quad \mathbf{R}_{\mathbf{nn}} = \begin{bmatrix} 4 & 0.5 & -0.1 \\ 0.5 & 1 & 0.2 \\ -0.1 & 0.2 & 2 \end{bmatrix},$$

$$\mathbf{R}_{\min} = \begin{bmatrix} 0.1 & -0.04 & 0 \\ -0.04 & 0.2 & 0 \\ 0 & 0 & 0.15 \end{bmatrix}, \quad \mathbf{R}_{\max} = \begin{bmatrix} 8 & 0.8 & 0 \\ 0.8 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

The experiment is carried in a similar manner as the experiment with $K = 2$ except that we vary the true $\mathbf{R}_{\boldsymbol{\theta\theta}}$ along the straight line between the two bounds (i.e., $\mathbf{R}_{\boldsymbol{\theta\theta}}(\lambda) = \lambda\mathbf{R}_{\min} + (1 - \lambda)\mathbf{R}_{\max}, 0 \leq \lambda \leq 1$) just for simplicity. The results are shown in Fig. 9. Once again, it is observed that while the conditional Γ-minimax estimate is optimal for the worst-case scenario in terms of conditional risk, it is evident that it still maintains competitive performance, in terms of MSE and frequentist risk, when compared with other methods.

Acknowledgment

The authors acknowledge a valuable discussion with Dr. Senjian An from which the proof of Lemma 3 could be completed. The authors also thank Dr. R. Howard for proofreading the manuscript.

References

[1] F. Alizadeh, "Interior point method in semidefinite programming with applications to combinatorial optimization," *SIAM J. Optimiz.*, vol. 5, pp. 13–51, 1995.
[2] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimation," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3145–3157, Sep. 2007.
[3] J. Berger, *Statistical Decision and Bayesian Analysis*. New York: Springer-Verlag, 1985.
[4] B. Betrò and F. Ruggeri, "Conditional Γ-minimax actions under convex losses," *Commun. Statist., Theory and Methods*, vol. 21, no. 4, pp. 1051–1066, 1992.
[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, MA: Cambridge University Press, 2004.
[6] L. D. Brown and R. C. Liu, "Bounds on the Bayes and minimax risk for signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 4–4, Jul. 1993.
[7] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*. London, U.K.: Chapman and Hall, 1996.
[8] R. M. Castro and R. D. Nowak, "Minimax bounds for active learning," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2339–2353, May 2008.
[9] A. DasGupta and W. Studden, "Frequentist behaviour of robust bayes estimates of normal means," in *Statist. and Decis.*, 1989, vol. 7, pp. 333–361 [Online]. Available: http://www.stat.purdue.edu/research/technical_reports/pdfs/1988/tr88-36c.pdf
[10] Y. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random variables," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 7–7, Jul. 2004.
[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore /London: Johns Hopkins Univ. Press, 1989.
[12] G. R. Lanckriet, L. E. Ghaoui, C. Bhattacharyya, and M. I. Jordan, "A robust minimax approach to classification," *J. Mach. Learn. Res.*, vol. 3, pp. 555–582, Dec. 2002.
[13] J. Lasserre, "A trace inequality for matrix product," *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1500–1501, Aug. 1995.
[14] E. E. Leamer, "Set of posterior means with bounded variance priors," *Econometrica*, vol. 50, no. 3, pp. 725–736, May 1982.
[15] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.
[16] J. Löfberg, "Yalmip: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004 [Online]. Available: http://control.ee.ethz.ch/~joloef/yalmip.php
[17] C. Morris, "Parametric empirical Bayes inference: Theory and applications," *J. Amer. Statist. Assoc.*, vol. 78, pp. 47–65, 1983.
[18] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.
[19] E. Polak, *Optimization: Algorithms and Consistent Approximations*. New York: Springer-Verlag, 1997.
[20] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statist. Applic. in Genetics and Molec. Biol.*, vol. 4, no. 1, 2005.
[21] L. L. Scharf, *Statistical Signal Processing: Detection Estimation and Time Series Analysis*. New York: Addison–Wesley, 1991.

[22] J. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimiz. Methods and Software*, vol. 11–12, pp. 625–653, 1999.

[23] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.

[24] B. Vidakovic, "Γ-minimax: A paradigm for conservative robust Bayesians," *Lecture Notes in Statistics*, vol. 152, pp. 241–259, 2000.

[25] S. R. Watson, "On Bayesian inference with incompletely specified prior distributions," *Biometrika*, vol. 61, no. 1, pp. 193–196, 1974.

Duc-Son Pham (S'02–M'05) received the Ph.D. in electrical engineering from Curtin University, Perth, Western Australia, in 2005.

He is a Research Fellow with the Department of Computing, Curtin University of Technology. His research interests include robust and scalable statistical methods for signal processing and computer vision. He has been a regular reviewer for numerous IEEE conferences and journals.

Hung H. Bui received the Ph.D. degree in computer science from Curtin University of Technology, Perth, Western Australia, in 1998.

He is a Senior Computer Scientist at the AI Center, SRI International, Menlo Park, CA. His main research interests are in the area of probabilistic inference, graphical models, and machine learning, especially models for sequential data with applications in activity recognition and natural language understanding. Prior to joining SRI, he was an assistant professor at the Department of Computing, Curtin University of Technology. During his high school years, he won a silver medal at the 1989 International Mathematics Olympiad in Braunschweig, Germany.

Svetha Venkatesh (SM'92) is John Curtin Distinguished Professor at the Department of Computing, Curtin University of Technology, Perth, Western Australia. He has extensive experience in low-level vision, pattern recognition and multimedia content analysis. Her research is in the areas of large scale pattern recognition, image understanding, multimedia analysis and applications of computer vision to image and video indexing and retrieval. She is the author of about 320 research papers in these areas and is currently the Director of the Institute of Multi-sensor Processing and Content Analysis. She is also the Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA, and *ACM Transactions on Multimedia Computing, Communications and Applications* (2008 onwards).

Dr. Venkatesh has been awarded the Fellowship of the International Association of Pattern Recognition and Fellow of Australian Academy of Technological Science and Engineering.