SIAM J. OPTIM.
Vol. 19, No. 3, pp. 1344–1367

A MINIMAX THEOREM WITH APPLICATIONS TO MACHINE LEARNING, SIGNAL PROCESSING, AND FINANCE*

SEUNG-JEAN KIM† AND STEPHEN BOYD†

Abstract. This paper concerns a fractional function of the form $x^T a/\sqrt{x^T Bx}$, where B is positive definite. We consider the game of choosing x from a convex set, to maximize the function, and choosing (a, B) from a convex set, to minimize it. We prove the existence of a saddle point and describe an efficient method, based on convex optimization, for computing it. We describe applications in machine learning (robust Fisher linear discriminant analysis), signal processing (robust beamforming and robust matched filtering), and finance (robust portfolio selection). In these applications, x corresponds to some design variables to be chosen, and the pair (a, B) corresponds to the statistical model, which is uncertain.

Key words. convex optimization, minimax theorem, robust optimization

AMS subject classifications. 65K0, 90C25, 90C47, 91B28

DOI. 10.1137/060677586

1. Introduction.

This paper concerns a fractional function of the form

$$f(x, a, B) = \frac{x^T a}{\sqrt{x^T Bx}}, \tag{1}$$

where $x, a \in \mathbb{R}^n$ and $B = B^T \in \mathbb{R}^{n\times n}$. We assume that $x \in \mathcal{X} \subseteq \mathbb{R}^n\backslash\{0\}$ and $(a, B) \in \mathcal{U} \subseteq \mathbb{R}^n \times \mathbb{S}^n_{++}$. Here $\mathbb{S}^n_{++}$ denotes the set of $n\times n$ symmetric positive definite matrices.

We list some of the basic properties of the function f. It is (positive) homogeneous (of degree zero) in x: for all $t > 0$,

$$f(tx, a, B) = f(x, a, B).$$

If

$$x^T a \geq 0 \text{ for all } x \in \mathcal{X} \text{ and for all } a, \text{ with } (a, B) \in \mathcal{U}, \tag{2}$$

then for fixed $(a, B) \in \mathcal{U}$, f is quasi-concave in x, and for fixed $x \in \mathcal{X}$, f is quasi-convex in (a, B). This can be seen as follows: for $\gamma \geq 0$, the set

$$\{x \mid f(a, B, x) \geq \gamma\} = \left\{x \;\middle|\; \gamma\sqrt{x^T Bx} \leq x^T a\right\}$$

is convex (since it is a second-order cone in $\mathbb{R}^n$), and the set

$$\{(a, B) \mid f(a, B, x) \leq \gamma\} = \left\{(a, B) \;\middle|\; \gamma\sqrt{x^T Bx} \geq x^T a\right\}$$

is convex (since $\sqrt{x^T Bx}$ is concave in B).

*Received by the editors December 12, 2006; accepted for publication (in revised form) May 6, 2008; published electronically November 21, 2008. This work was funded in part by the Precourt Institute on Energy Efficiency, Army award W911NF-07-1-0029, NSF award 0529426, NASA award NNX07AEIIA, AFOSR award FA9550-06-1-0514, and AFOSR award FA9550-06-1-0312.
http://www.siam.org/journals/siopt/19-3/67758.html

†Information Systems Laboratory, Electrical Engineering Department, Stanford University, Stanford, CA 94305-9510 (sjkim@stanford.edu, boyd@stanford.edu).

A zero-sum game and related problems. In this paper we consider the zero-sum game of choosing x from a convex set $\mathcal{X}$, to maximize the function, and choosing (a, B) from a convex compact set $\mathcal{U}$, to minimize it. The game is associated with the following two problems:

- max-min problem

$$\begin{array}{ll} \text{maximize} & \inf_{(a,B)\in\mathcal{U}} f(x,a,B) \\ \text{subject to} & x \in \mathcal{X}, \end{array} \tag{3}$$

with variables $x \in \mathbb{R}^n$,

- min-max problem

$$\begin{array}{ll} \text{minimize} & \sup_{x\in\mathcal{X}} f(x,a,B) \\ \text{subject to} & (a,B) \in \mathcal{U}, \end{array} \tag{4}$$

with variables $a \in \mathbb{R}^n$ and $B = B^T \in \mathbb{R}^{n\times n}$.

Problems of the form (3) arise in several disciplines including machine learning (robust Fisher linear discriminant analysis), signal processing (robust beamforming and robust matched filtering), and finance (robust portfolio selection). In these applications, x corresponds to some design variables to be chosen, and the pair (a, B) corresponds to the first and second moments of a random vector, say, $\mathbf{Z}$, which are uncertain. We want to choose x so that the combined random variable $x^T\mathbf{Z}$ is well separated from zero. The ratio of the mean of the random variable to the standard deviation $f(x, a, B)$ measures the extent to which the random variable can be well separated from zero. The max-min problem is to find the design variables that are optimal in a worst-case sense, where worst-case means over all possible statistics. The min-max problem is to find the least-favorable statistical model, with the design variables chosen optimally for the statistics.

Minimax properties. The minimax inequality or *weak minimax property*

$$\sup_{x\in\mathcal{X}} \inf_{(a,B)\in\mathcal{U}} f(x,a,B) \le \inf_{(a,B)\in\mathcal{U}} \sup_{x\in\mathcal{X}} f(x,a,B) \tag{5}$$

always holds for any $\mathcal{X} \subseteq \mathbb{R}$ and any $\mathcal{U} \subseteq \mathbb{S}^n_{++}$. The minimax equality or *strong minimax property*

$$\sup_{x\in\mathcal{X}} \inf_{(a,B)\in\mathcal{U}} f(x,a,B) = \inf_{(a,B)\in\mathcal{U}} \sup_{x\in\mathcal{X}} f(x,a,B) \tag{6}$$

holds if $\mathcal{X}$ is convex, $\mathcal{U}$ is convex and compact, and (2) holds, which follows from Sion's quasi-convex–quasi-concave minimax theorem [25].

In this paper we will show that the strong minimax property holds with a weaker assumption than (2):

$$\text{there exists } \bar{x} \in \mathcal{X} \text{ such that } \bar{x}^T a > 0 \text{ for all } a \text{ with } (a,B) \in \mathcal{U}. \tag{7}$$

To state the minimax result, we first describe an equivalent formulation of the min-max problem (4).

PROPOSITION 1. *Suppose that $\mathcal{X}$ is a cone in $\mathbb{R}^n$ that does not contain the origin, with $\mathcal{X} \cup \{0\}$ convex and closed, and $\mathcal{U}$ is a compact subset of $\mathbb{R}^n \times \mathbb{S}^n_{++}$. Suppose further that* (7) *holds. Then, the min-max problem* (4) *is equivalent to*

$$\begin{array}{ll} \textit{minimize} & (a+\lambda)^T B^{-1} (a+\lambda) \\ \textit{subject to} & (a,B) \in \mathcal{U}, \quad \lambda \in \mathcal{X}^*, \end{array} \tag{8}$$

where $a \in \mathbb{R}^n$, $B = B^T \in \mathbb{R}^{n\times n}$, and $\lambda \in \mathbb{R}^n$ are the variables and $\mathcal{X}^*$ is the dual *cone of $\mathcal{X}$ given by*

$$\mathcal{X}^* = \left\{\lambda \in \mathbb{R}^n \mid \lambda^T x \geq 0 \ \forall x \in \mathcal{X}\right\},$$

in the following sense: if $(a^\star, B^\star, \lambda^\star)$ solves (8), then $(a^\star, B^\star)$ solves (4), and conversely if $(a^\star, B^\star)$ solves (4), then there exists $\lambda^\star \in \mathcal{X}^*$ such that $(a^\star, B^\star, \lambda^\star)$ solves (8). Moreover,

$$\inf_{(a,B)\in\mathcal{U}} \sup_{x\in\mathcal{X}} f(x,a,B) = \left(\inf_{(a,B)\in\mathcal{U},\ \lambda\in\mathcal{X}^*} (a+\lambda)^T B^{-1}(a+\lambda)\right)^{1/2}.$$

Finally, (8) always has a solution, and for any solution $(a^\star, B^\star, \lambda^\star)$,

$$a^\star + \lambda^\star \neq 0.$$

The proof is deferred to the appendix.

The dual cone $\mathcal{X}^*$ is always convex. The objective of (8) is convex since a function of the form $f(x, X) = x^T X^{-1} x$, called a matrix fractional function, is convex over $\mathbb{R}^n \times \mathbb{S}^n_{++}$; see, e.g., [7, section 3.1.7]. Therefore, (8) is a convex problem. We conclude that the min-max problem (4) can be reformulated as the convex problem (8).

We can solve the max-min problem (3), using a minimax result for the fractional function $f(x, a, B)$.

THEOREM 1. *Suppose that $\mathcal{X}$ is a cone in $\mathbb{R}^n$ that does not contain the origin, with $\mathcal{X} \cup \{0\}$ convex and closed, and $\mathcal{U}$ is a convex compact subset of $\mathbb{R}^n \times \mathbb{S}^n_{++}$. Suppose further that (7) holds. Let $(a^\star, B^\star, \lambda^\star)$ be a solution to the convex problem (8) (whose existence is guaranteed in Proposition 1). Then,*

$$x^\star = B^{\star -1}(a^\star + \lambda^\star) \in \mathcal{X},$$

and the triple $(x^\star, a^\star, B^\star)$ satisfies the saddle-point property

$$f(x, a^\star, B^\star) \leq f(x^\star, a^\star, B^\star) \leq f(x^\star, a, B) \quad \forall x \in \mathcal{X} \quad \forall (a, B) \in \mathcal{U}. \tag{9}$$

The proof is deferred to the appendix.

We show that the assumption (7) is needed for the strong minimax property to hold. Consider $\mathcal{X} = \mathbb{R}^n \backslash \{0\}$ and $\mathcal{U} = \mathcal{B}_1 \times \{I\}$, where $\mathcal{B}_1$ is the Euclidean ball of radius one. Then, all of the assumptions hold except for (7). We have

$$\sup_{x\neq 0} \inf_{(a,B)\in\mathcal{U}} \frac{x^T a}{\sqrt{x^T B x}} = \sup_{x\neq 0} \inf_{a\in\mathcal{B}_1} \frac{x^T a}{\sqrt{x^T x}} = \sup_{x\neq 0} \frac{-\|x\|}{\|x\|} = -1$$

and

$$\inf_{(a,B)\in\mathcal{U}} \sup_{x\neq 0} \frac{x^T a}{\sqrt{x^T B x}} = \inf_{a\in\mathcal{B}_1} \sup_{x\neq 0} \frac{x^T a}{\|x\|} = \inf_{a\in\mathcal{B}_1} \frac{\|x\|\|a\|}{\|x\|} = 0.$$

From a standard result [2, section 2.6] in minimax theory, the saddle-point property (9) means that

$$\begin{aligned}
f(x^\star, a^\star, B^\star) &= \sup_{x\in\mathcal{X}} f(x, a^\star, B^\star) \\
&= \inf_{(a,B)\in\mathcal{U}} f(x^\star, a, B) \\
&= \sup_{x\in\mathcal{X}} \inf_{(a,B)\in\mathcal{U}} f(x, a, B) \\
&= \inf_{(a,B)\in\mathcal{U}} \sup_{x\in\mathcal{X}} f(x, a, B).
\end{aligned}$$

As a consequence, $x^\star$ solves (3).

More computational results. The max-min problem (3) has a unique solution up to (positive) scaling.

PROPOSITION 2. *Under the assumptions of Theorem* 1, *the max-min problem* (3) *has a unique solution up to (positive) scaling, meaning that for any two solutions* $x^\star$ *and* $y^\star$, *there is a positive number* $\alpha > 0$ *such that* $x^\star = \alpha y^\star$.

The proof is deferred to the appendix.

The convex problem (8) can be reformulated as a standard convex optimization problem. Using the Schur complement technique [7, Appendix 5.5], we can see that

$$(a+\lambda)^T B^{-1}(a+\lambda) \le t$$

if and only if the linear matrix inequality (LMI)

$$\begin{bmatrix} t & (a+\lambda)^T \\ a+\lambda & B \end{bmatrix} \succeq 0$$

holds. (Here $A \succeq 0$ means that A is positive semidefinite.) The convex problem (8) is therefore equivalent to

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & (a,B) \in \mathcal{U}, \quad \lambda \in \mathcal{X}^*, \quad \begin{bmatrix} t & (a+\lambda)^T \\ a+\lambda & B \end{bmatrix} \succeq 0, \end{array}$$

where the variables are $t \in \mathbb{R}$, $a \in \mathbb{R}^n$, $B = B^T \in \mathbb{R}^{n\times n}$, and $\lambda \in \mathbb{R}^n$. When the uncertainty sets $\mathcal{U}$ can be represented by LMIs, this problem is a semidefinite program (SDP). (Several high-quality open-source solvers for SDPs are available, e.g., SeDuMi [26], SDPT3 [27], and DSDP5 [1].) The reader is referred to [6, 29] for more on semidefinite programming and LMIs.

Outline of the paper. In the next section, we give a probabilistic interpretation of the saddle-point property established above. In sections 3–5, we give the applications of the minimax result in machine learning, signal processing, and portfolio selection. We give our conclusions in section 6. The appendix contains the proofs that are omitted from the main text.

2. A probabilistic interpretation.

2.1. Probabilistic linear separation. Suppose $z \sim \mathcal{N}(a,B)$ and $x \in \mathbb{R}^n$. Here, we use $\mathcal{N}(a,B)$ to denote the Gaussian distribution with mean a and covariance B. Then, $x^Tz \sim \mathcal{N}(x^Ta, x^TBx)$, so

$$\textbf{Prob}\left(x^Tz \ge 0\right) = \Phi\left(\frac{x^Ta}{\sqrt{x^TBx}}\right), \tag{10}$$

where Φ is the cumulative distribution function of the standard normal distribution.

Theorem 1 with $\mathcal{U} = \{(a,B)\}$ tells us that the right-hand side of (10) is maximized (over $x \in \mathcal{X}$) by $x = B^{-1}(a+\lambda^\star)$, where $\lambda^\star$ solves the convex problem (8) with $\mathcal{U} = \{(a,B)\}$. In other words, $x = B^{-1}(a+\lambda^\star)$ gives the hyperplane through the origin that maximizes the probability of z being on its positive side. The associated maximum probability is $\Phi([(a+\lambda^\star)^T B^{-1}(a+\lambda^\star)]^{1/2})$. Thus, $(a+\lambda^\star)^T B^{-1}(a+\lambda^\star)$ (which is the objective of (8)) can be used to measure the extent to which a hyperplane perpendicular to $x \in \mathcal{X}$ can separate a random signal $z \sim \mathcal{N}(a,B)$ from the origin.

We give another interpretation. Suppose that we know the mean $\mathbf{E}\,z = a$ and the covariance $\mathbf{E}(z-a)(z-a)^T = B$ of z, but its third and higher moments are unknown.

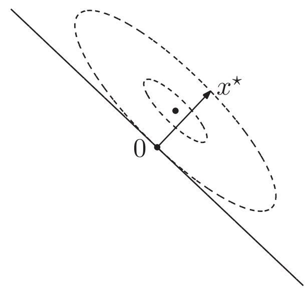

FIG. 1. *Illustration of $x^\star = B^{-1}a$. The center of the two confidence ellipsoids (whose boundaries are shown as dashed line curves) is a, and their shapes are determined by B.*

Here $\mathbf{E}$ denotes the expectation operation. Then, $\mathbf{E}\, x^T z = x^T a$ and $\mathbf{E}(x^T z - x^T a)^2 = x^T B x$, so by the Chebyshev bound, we have

$$\mathbf{Prob}\left(x^T z \geq 0\right) \geq \Psi\left(\frac{x^T a}{\sqrt{x^T B x}}\right), \tag{11}$$

where

$$\Psi(u) = \frac{\max\{u,0\}^2}{1+\max\{u,0\}^2}.$$

This bound is sharp; in other words, there is a distribution for z with mean a and covariance B for which equality holds in (11) [3, 30]. Since Ψ is increasing, this probability is also maximized by $x = B^{-1}(a+\lambda^\star)$. Thus, $x = B^{-1}(a+\lambda^\star)$ gives the hyperplane through the origin and perpendicular to $x \in \mathcal{X}$ that maximizes the Chebyshev lower bound for $\mathbf{Prob}(x^T z \geq 0)$. The maximum value of the Chebyshev lower bound is $p^\star/(1+p^\star)$, where $p^\star = \left[(a+\lambda^\star)^T B^{-1}(a+\lambda^\star)\right]^{1/2}$. This quantity assesses the maximum extent to which a hyperplane perpendicular to $x \in \mathcal{X}$ can separate from the origin a random signal z, whose first and second moments are known but otherwise arbitrary. This quantity is an increasing function of $p^\star$, so the hyperplane perpendicular to $x \in \mathcal{X}$ that maximally separates from the origin a Gaussian random signal $z \sim \mathcal{N}(a,B)$ also maximally separates, in the sense of the Chebyshev bound, a signal with known mean and covariance.

When $\mathcal{X} = \mathbb{R}^n \backslash \{0\}$, we have $\mathcal{X}^* = 0$, so $x = B^{-1}a$ maximizes the right-hand side of (10). We can give its graphical interpretation. We find the confidence ellipsoid of the Gaussian distribution $\mathcal{N}(a,B)$, whose boundary touches the origin. This ellipsoid is tangential to the hyperplane through the origin and perpendicular to $x = B^{-1}a$. Figure 1 illustrates this interpretation in $\mathbb{R}^2$.

2.2. Robust linear separation. We now assume that the mean and covariance are uncertain but known to belong to a convex compact subset $\mathcal{U}$ of $\mathbb{R}^n \times \mathbb{S}^n_{++}$. We make one assumption: for each $(a,\Sigma) \in \mathcal{U}$, we have $a \neq 0$. In other words, we rule out the possibility that the mean is zero.

Theorem 1 tells us that there exists a triple $(x^\star, a^\star, B^\star)$, with $x^\star \in \mathcal{X}$ and $(a^\star, B^\star) \in \mathcal{U}$, such that

$$\Phi\left(\frac{x^T a^\star}{\sqrt{x^T B^\star x}}\right) \leq \Phi\left(\frac{x^{\star T} a^\star}{\sqrt{x^{\star T} B^\star x^\star}}\right) \leq \Phi\left(\frac{x^{\star T} a}{\sqrt{x^{\star T} B x^\star}}\right) \quad \forall x \in \mathcal{X} \quad \forall (a,B) \in \mathcal{U}. \tag{12}$$

Here we use the fact that Φ is strictly increasing.

From the saddle-point property (12), we can see that $x^\star$ solves

$$\begin{array}{ll} \text{maximize} & \inf_{(a,B)\in\mathcal{U}, z\sim\mathcal{N}(a,B)} \mathbf{Prob}\left(x^T z \geq 0\right) \\ \text{subject to} & x \in \mathcal{X}, \end{array} \tag{13}$$

and the pair $(a^\star, B^\star)$ solves

$$\begin{array}{ll} \text{minimize} & \sup_{x\in\mathcal{X}, z\sim\mathcal{N}(a,B)} \mathbf{Prob}\left(x^T z > 0\right) \\ \text{subject to} & (a,B) \in \mathcal{U}. \end{array} \tag{14}$$

Problem (13) is to find a hyperplane through the origin and perpendicular to $x \in \mathcal{X}$ that separates robustly a normal random variable z on $\mathbb{R}^n$ with uncertain first and second moments belonging to $\mathcal{U}$. Problem (14) is to find the least-favorable model in terms of the separation probability (when the random variable is normal). It follows from (10) that (13) is equivalent to the max-min problem (3), and (14) is equivalent to (4) and hence to the convex problem (8) by Proposition 1. These two problems can be solved using convex optimization.

We close by pointing out that the same results hold with the Chebyshev bound as the separation probability.

3. Robust Fisher discriminant analysis.

As another application, we consider a robust classification problem.

3.1. Fisher linear discriminant analysis.

In linear discriminant analysis, we want to separate two classes which can be identified with two random variables in $\mathbb{R}^n$. Fisher linear discriminant analysis (FLDA) is a widely used technique for pattern classification, proposed by Fisher in the 1930s. The reader is referred to standard textbooks on statistical learning, e.g., [13], for more on FLDA.

For a (linear) discriminant characterized by $w \in \mathbb{R}^n$, the degree of discrimination is measured by the Fisher discriminant ratio

$$F(w, \mu_+, \mu_-, \Sigma_+, \Sigma_-) = \frac{\left(w^T(\mu_+ - \mu_-)\right)^2}{w^T(\Sigma_+ + \Sigma_-)w},$$

where μ_+ and Σ_+ (μ_+ and Σ_-) denote the mean and covariance, respectively, of examples drawn from the positive (negative) class. A discriminant that maximizes the Fisher discriminant ratio is given by

$$\bar{w} = (\Sigma_+ + \Sigma_-)^{-1}(\mu_+ - \mu_-),$$

which gives the maximum Fisher discriminant ratio

$$\sup_{w\neq 0} F(w, \mu_+, \mu_-, \Sigma_+, \Sigma_-) = (\mu_+ - \mu_-)^T(\Sigma_+ + \Sigma_-)^{-1}(\mu_+ - \mu_-).$$

Once the optimal discriminant is found, we can form the (binary) classifier

$$\phi(x) = \operatorname{sgn}\left(\bar{w}^T x + v\right), \tag{15}$$

where

$$\operatorname{sgn}(z) = \begin{cases} +1, & z > 0, \\ -1, & z \leq 0, \end{cases}$$

and v is the bias or threshold. The classifier picks the outcome, given x, according to the linear boundary between the two binary outcomes (defined by $\bar{w}^T x + v = 0$).

We can give a probabilistic interpretation of FLDA. Suppose that $x \sim \mathcal{N}(\mu_+, \Sigma_+)$ and $y \sim \mathcal{N}(\mu_-, \Sigma_-)$. We want to find w that maximizes $\mathbf{Prob}(w^T x > w^T y)$. Here,

$$x - y \sim \mathcal{N}(\mu_+ - \mu_-, \Sigma_+ + \Sigma_-),$$

so

$$\mathbf{Prob}\left(w^T x > w^T y\right) = \mathbf{Prob}\left(w^T (x - y) > 0\right) = \Phi\left(\frac{w^T(\mu_+ - \mu_-)}{\sqrt{w^T(\Sigma_+ + \Sigma_-)w}}\right).$$

This probability is called the nominal discrimination probability. Evidently, FLDA amounts to maximizing the fractional function

$$f(w, \mu_+ - \mu_-, \Sigma_+ + \Sigma_-) = \frac{w^T(\mu_+ - \mu_-)}{\sqrt{w^T(\Sigma_+ + \Sigma_-)w}}.$$

3.2. Robust Fisher linear discriminant analysis. In FLDA, the problem data or parameters (i.e., the first and second moments of the two random variables) are not known but are estimated from sample data. FLDA can be sensitive to the variation or uncertainty in the problem data, meaning that the discriminant computed from an estimate of the parameters can give very poor discrimination for another set of problem data that is also a reasonable estimate of the parameters. Robust FLDA attempts to systematically alleviate this sensitivity problem by explicitly incorporating a model of data uncertainty in the classification problem and optimizing for the worst-case scenario under this model; see [17] for more on robust FLDA and its extension.

We assume that the problem data μ_+, μ_-, Σ_+, and Σ_- are uncertain but known to belong to a convex compact subset $\mathcal{V}$ of $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{S}^n_{++} \times \mathbb{S}^n_{++}$. We make the following assumption:

$$\text{for each } (\mu_+, \mu_-, \Sigma_+, \Sigma_-) \in \mathcal{V}, \text{ we have } \mu_+ \neq \mu_-. \tag{16}$$

This assumption simply means that for each possible value of the means and covariances, the two classes are distinguishable via FLDA.

The *worst-case analysis problem* of finding the worst-case means and covariances for a given discriminant w can be written as

$$\begin{array}{ll} \text{minimize} & f(w, \mu_+ - \mu_-, \Sigma_+ + \Sigma_-) \\ \text{subject to} & (\mu_+, \mu_-, \Sigma_+, \Sigma_-) \in \mathcal{V}, \end{array} \tag{17}$$

with variables μ_+, μ_-, Σ_+, and Σ_-. Optimal points for this problem, say, $(\mu_+^{\text{wc}}, \mu_-^{\text{wc}}, \Sigma_+^{\text{wc}}, \Sigma_-^{\text{wc}})$, are called the *worst-case means and covariances*, which depend on w. With the worst-case means and covariances, we can compute the *worst-case discrimination probability*

$$\mathbf{P}_{\text{wc}}(w) = \Phi\left(\frac{w^T(\mu_+^{\text{wc}} - \mu_-^{\text{wc}})}{\sqrt{w^T(\Sigma_+^{\text{wc}} + \Sigma_-^{\text{wc}})w}}\right)$$

(over the set $\mathcal{U}$ of possible means and covariances).

The *robust FLDA problem* is to find a discriminant that maximizes the worst-case Fisher discriminant ratio:

$$\begin{array}{ll} \text{maximize} & \inf_{(\mu_+,\mu_-,\Sigma_+,\Sigma_-)\in\mathcal{V}} f(w, \mu_+ - \mu_-, \Sigma_+ + \Sigma_-) \\ \text{subject to} & w \neq 0, \end{array} \tag{18}$$

with variable w. Here we choose a linear discriminant that maximizes the Fisher discrimination ratio, with the worst possible means and covariances that are consistent with our data uncertainty model. Any solution to (18) is called a *robust optimal Fisher discriminant*.

The robust FLDA problem (18) has the form (3) with

$$\mathcal{U} = \{(\mu_+ - \mu_-, \Sigma_+ + \Sigma_-) \in \mathbb{R}^n \times \mathbb{S}^n_{++} \mid (\mu_+, \mu_-, \Sigma_+, \Sigma_-) \in \mathcal{U}\}.$$

In this problem, each element of the set $\mathcal{U}$ is a pair of the mean and covariance of the difference of the two random variables. For this problem, we can see from (16) that assumption (7) holds. The robust FLDA problem can therefore be solved by using the minimax result described above.

3.3. Numerical example. We illustrate the result with a classification problem in $\mathbb{R}^2$. The nominal means and covariances of the two classes are

$$\bar{\mu}_+ = (1, 0), \qquad \bar{\mu}_- = (-1, 0), \qquad \bar{\Sigma}_+ = \bar{\Sigma}_- = I \in \mathbb{R}^{2\times 2}.$$

We assume that only μ_+ is uncertain and lies within the ellipse

$$\mathcal{E} = \left\{\mu_+ \in \mathbb{R}^2 \mid \mu_+ = \bar{\mu}_+ + Pu,\ \|u\| \leq 1\right\},$$

where the matrix P which determines the shape of the ellipse is

$$P = \begin{bmatrix} 0.78 & 0.64 \\ 0.64 & 0.78 \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

Figure 2 illustrates the setting described above. Here the shaded ellipse corresponds to $\mathcal{E}$, and the dashed line curves are the set of points μ_+ and μ_- that satisfy

$$\left\|{\Sigma_+}^{-1/2}(\mu_+ - \bar{\mu}_+)\right\| = \|\mu_+ - \bar{\mu}_+\| = 1, \qquad \left\|{\Sigma_-}^{-1/2}(\mu_- - \bar{\mu}_-)\right\| = \|\mu_- - \bar{\mu}_-\| = 1.$$

The nominal optimal discriminant which maximizes the Fisher discriminant ratio with the nominal means and covariances is given by $w^{\text{nom}} = (1, 0)$. The robust optimal discriminant w^{rob} is computed using the method described above. Figure 2 shows two linear decision boundaries

$$x^T w^{\text{nom}} = 0, \qquad x^T w^{\text{rob}} = 0$$

determined by the two discriminants. Since the mean of the positive class is uncertain and the uncertainty is significant in a certain direction, the robust discriminant is tilted toward the direction.

Table 1 summarizes the results. Here, $\mathbf{P}_{\text{nom}}$ is the nominal discrimination probability and $\mathbf{P}_{\text{wc}}$ is the worst-case discrimination probability. The nominal optimal discriminant achieves $\mathbf{P}_{\text{nom}} = 0.92$, which corresponds to 92% of correct discrimination without uncertainty. However, with uncertainty present, its nominal discrimination probability degrades rapidly; the worst-case discrimination probability for the nominal optimal discriminant is 78%. The robust optimal discriminant performs well in the presence of uncertainty. It has a worst-case discrimination probability around 83%, 5% higher than that of the nominal optimal discriminant.

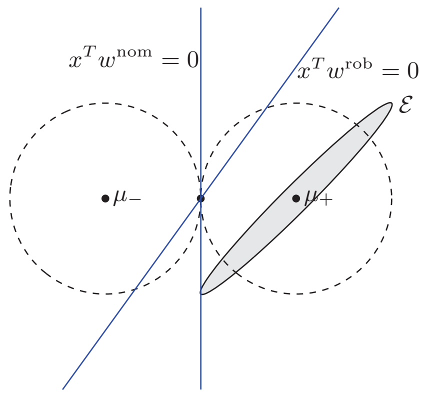

FIG. 2. *A simple example for robust FLDA.*

TABLE 1
Robust discriminant analysis results.

	$\mathbf{P}_{\mathrm{nom}}$	$\mathbf{P}_{\mathrm{wc}}$
Nominal optimal discriminant	0.92	0.78
Robust optimal discriminant	0.87	0.83

4. Robust matched filtering. We consider a signal model of the form

$$y(t) = s(t)a + v(t) \in \mathbb{R}^n,$$

where a is the steering vector, $s(t) \in \{0, 1\}$ is the binary source signal, $y(t) \in \mathbb{R}^n$ is the received signal, and $v(t) \sim \mathcal{N}(0, \Sigma)$ is the noise. We consider the problem of estimating $s(t)$, based on an observed sample of y. In other words, the sample is generated from one of the two possible distributions $\mathcal{N}(0, \Sigma)$ and $\mathcal{N}(a, \Sigma)$, and we are to guess which one.

After reviewing a basic result on optimal detection with the setting described above, we show how the minimax result given above allows us to design a robust detector that takes into account the uncertainty in the model parameters, namely, the steering vector and the noise covariance.

4.1. Matched filtering. A (deterministic) detector is a function ψ from $\mathbb{R}^n$ (the set of possible observed values) into $\{0, 1\}$ (the set of possible signal values or hypotheses). It can be expressed as

$$\psi(y) = \begin{cases} 0, & h(y) < t, \\ 1, & h(y) > t, \end{cases} \tag{19}$$

which thresholds a detection or test statistic, a function of the received signal, $h(y) \in \mathbb{R}$. Here t is the threshold that determines the boundary between the two hypotheses. A detector with a detection statistic of the form $h(y) = w^T y$ is called linear.

The performance of a detector ψ can be summarized by the pair $(P_{\mathrm{fp}}, P_{\mathrm{tp}})$, where

$$P_{\mathrm{fp}} = \mathbf{Prob}(\psi(y) = 1 \mid s(t) = 0)$$

is the *false positive or alarm rate* (the probability that the signal is falsely detected when in fact there is no signal) and

$$P_{\mathrm{tp}} = \mathbf{Prob}(\psi(y) = 1 \mid s(t) = 1)$$

is the *true positive rate* (the probability that the signal is detected correctly). The optimal detector design problem is a bicriterion problem, with objectives P_{fn} and P_{fp}. The optimal trade-off curve between P_{fn} and P_{fp} is called the *receiver operating characteristic* (ROC).

The filtered output, with weight vector $w \in \mathbb{R}^n$, is given by

$$w^T y(t) = s(t) w^T a + w^T v(t).$$

The power of the steering vector $w^T a$ (which is deterministic) at the filtered output is given by $(w^T a)^2$, and the power of the undesired signal $w^T v$ at the filtered output is $w^T \Sigma w$. The signal to noise ratio (SNR) is

$$S(w, a, \Sigma) = \frac{\left(w^T a\right)^2}{w^T \Sigma w}.$$

The optimal ROC curve is obtained using a linear detection statistic $h(y) = w^{\star T} y$ with $w^\star$ maximizing

$$f(w, a, \Sigma) = \frac{w^T a}{\sqrt{w^T \Sigma w}},$$

which is the square root of the SNR (SSNR). (See, e.g., [28].) The weight vector that maximizes SSNR is given by $w = \Sigma^{-1} a$. When the covariance is a scaled identity matrix, the matched filter $w = a$ is optimal. Even when Σ is not a scaled identity matrix, the optimal weight vector is called the matched filter.

4.2. Robust matched filtering. Matched filtering is often sensitive to the uncertainty in the input parameters, namely, the steering vector and the noise covariance. Robust matched filtering attempts to alleviate the sensitivity problem by taking into account an uncertainty model in the detection problem. (The reader is referred to the tutorial [15] for more on robust signal detection.)

We assume that the desired signal and covariance matrix are uncertain but known to belong to a convex compact subset $\mathcal{U}$ of $\mathbb{R}^n \times \mathbb{S}^n_{++}$. We make a technical assumption:

$$a \neq 0 \qquad \forall (a, \Sigma) \in \mathcal{U}. \tag{20}$$

In other words, we rule out the possibility that the signal we want to detect is zero.

The *worst-case SSNR analysis problem* of finding a steering vector and a covariance that minimize SSNR for a given weight vector w can be written as

$$\begin{array}{ll} \text{minimize} & f(w, a, \Sigma) \\ \text{subject to} & (a, \Sigma) \in \mathcal{U}, \end{array} \tag{21}$$

with variables a and Σ. The optimal value of this problem is the *worst-case SSNR* (over the uncertainty set $\mathcal{U}$).

The *robust matched-filtering problem* is to find a weight vector that maximizes the worst-case SSNR, which can be cast as

$$\begin{array}{ll} \text{maximize} & \inf\limits_{(a,\Sigma) \in \mathcal{U}} f(x, a, B) \\ \text{subject to} & w \neq 0, \end{array} \tag{22}$$

with variables w. (The thresholding rule $h(y) = w^{\star T} y$ that uses a solution $w^\star$ of this problem as the weight vector yields the robust ROC curve that characterizes limits of performance in the worst-case sense.)

The robust signal detection setting described above is exactly the minimax setting described in the introduction, where a is the steering vector and B is the noise covariance. For this problem, we can see from the compactness of $\mathcal{U}$ and (20) that assumption (7) holds. We can solve the robust matched-filtering problem (22), using the minimax result for the fractional function (1).

We close by pointing out that we can handle convex constraints on the weight vector. For example, in robust beamforming, a special type of robust matched-filtering problem, we often want to choose the weight vector that maximizes the worst-case SSNR, subject to a unit array gain for the desired wave and rejection constraints on interferences [22]. This problem can also be solved using Theorem 1.

4.3. Numerical example. As an illustrative example, we consider the case when $a = (2,3,2,2)$ is fixed (with no uncertainty) and the noise covariance Σ is uncertain and has the form

$$\begin{bmatrix} 1 & - & + & - \\ & 1 & ? & + \\ & & 1 & ? \\ & & & 1 \end{bmatrix}.$$

(Only the upper triangular part is shown because the matrix is symmetric.) Here, "+" means that $\Sigma_{ij} \in [0,1]$, "−" means that $\Sigma_{ij} \in [-1,0]$, and "?" means that $\Sigma_{ij} \in [-1,1]$. Of course we assume $\Sigma \succ 0$. The nominal noise covariance is taken as

$$\bar{\Sigma} = \begin{bmatrix} 1 & -.5 & .5 & -.5 \\ & 1.0 & 0.0 & .5 \\ & & 1.0 & 0.0 \\ & & & 1.0 \end{bmatrix}.$$

Here, the upper-triangular part is shown since the matrix is symmetric. With the nominal covariance, we compute the nominal optimal weight vector or filter.

The least-favorable covariance, found by solving the convex problem (8) corresponding to the problem data above, is given by

$$\Sigma^{\mathrm{lf}} = \begin{bmatrix} 1.00 & 0.00 & .38 & -.12 \\ & 1.00 & .41 & .74 \\ & & 1.00 & .23 \\ & & & 1.00 \end{bmatrix}.$$

With the least-favorable covariance, we compute the robust optimal weight vector or filter.

Table 2 summarizes the results. The nominal optimal filter achieves an SSNR of 5.5 without uncertainty. In the presence of uncertainty, the SSNR achieved by the filter can degrade rapidly; the worst-case SSNR level for the nominal optimal filter is 3.0. The robust filter performs well in the presence of model mismatch; it has the worst-case SSNR of 3.6, which is 20% larger than that of the nominal optimal filter.

5. Worst-case Sharpe ratio maximization. The minimax result has an important application in robust portfolio selection.

5.1. Mean-variance asset allocation. Since the pioneering work of Markowitz [20], mean-variance (MV) analysis has been a topic of extensive research. In MV analysis, the (percentage) returns of risky assets $1,\ldots,n$ over a period are modeled

TABLE 2
Robust matched-filtering results.

	Nominal SSNR	Worst-case SSNR
Nominal optimal filter	5.5	3.0
Robust optimal filter	4.9	3.6

as a random vector $a = (a_1, \ldots, a_n)$ in $\mathbb{R}^n$. The input data or parameters needed for MV analysis are the mean μ and the covariance matrix Σ of a:

$$\mu = \mathbf{E}\, a, \qquad \Sigma = \mathbf{E}\, (a-\mu)(a-\mu)^T.$$

We assume that there is a risk-free asset with deterministic return μ_{rf} and zero variance.

A portfolio $w \in \mathbb{R}^{n+1}$ is a finite linear combination of the assets. Let w_i denote the amount of asset i held throughout the period. A long position in asset i corresponds to $w_i > 0$, and a short position in asset i corresponds to $w_i < 0$. The return of a portfolio $w = (w_1, \ldots, w_n)$ is a (scalar) random variable $w^T a = \sum_{i=1}^n w_i a_i$, whose mean and volatility (standard deviation) are $\mu^T w$ and $\sqrt{w^T \Sigma w}$, respectively. We assume that an admissible portfolio $w = (w_1, \ldots, w_n)$ is constrained to lie in a convex compact subset $\mathcal{A}$ of $\mathbb{R}^n$. The portfolio budget constraint on w can be expressed, without loss of generality, as $\mathbf{1}^T w = 1$. Here $\mathbf{1}$ is the vector of all ones. The set of admissible portfolios subject to the portfolio budget constraint is given by

$$\mathcal{W} = \left\{ w \mid w \in \mathcal{A},\ \mathbf{1}^T w = 1 \right\}.$$

The performance of an admissible portfolio is often measured by its reward-to-variability or Sharpe ratio (SR):

$$S(w, \mu, \Sigma) = \frac{\mu^T w - \mu_{\mathrm{rf}}}{\sqrt{w^T \Sigma w}}.$$

The admissible portfolio that maximizes the ratio over $\mathcal{W}$ is called the *tangency portfolio* (TP). The SR achieved by this portfolio is called the market price of risk. The TP plays an important role in asset pricing theory and practice (see, e.g., [8, 19, 24]).

If the n risky assets with (single period) returns follow $a \sim \mathcal{N}(\mu, \Sigma)$, then

$$w^T a \sim \mathcal{N}\left(w^T \mu, w^T \Sigma w\right),$$

so the probability of outperforming the risk-free return μ_{rf} is

$$\mathbf{Prob}\left(a^T w > \mu_{\mathrm{rf}}\right) = \Phi\left(\frac{\mu^T w - \mu_{\mathrm{rf}}}{\sqrt{w^T \Sigma w}}\right).$$

This probability is maximized by the TP.

SR maximization is related to the safety-first approach to portfolio selection [23], through the Chebyshev bound. Suppose $\mathbf{E}\, a = \mu$, $\mathbf{E}\,(a-\mu)^T(a-\mu) = \Sigma$ and otherwise arbitrary. Then, $\mathbf{E}\, a^T w = \mu^T x$ and $\mathbf{E}\,(a^T x - \mu^T x)^2 = w^T \Sigma w$, so it follows from the Chebyshev bound that

$$\mathbf{Prob}\left(a^T w \geq \mu_{\mathrm{rf}}\right) \geq \Psi\left(\frac{\mu^T w - \mu_{\mathrm{rf}}}{\sqrt{w^T \Sigma w}}\right).$$

In the safety-first approach [23], we want to find a portfolio that maximizes the bound. Since Ψ is increasing, this bound is also maximized by the TP.

5.2. Worst-case SR maximization. The input parameters are estimated with error. Conventional MV allocation is often sensitive to the uncertainty or the estimation error in the parameters, meaning that optimal portfolios computed with an estimate of the parameters can give very poor performance for another set of parameters that is similar and statistically hard to distinguish from the estimate; see, e.g., [4, 5, 14, 21], to name a few. Robust MV portfolio analysis attempts to systematically alleviate the sensitivity problem of conventional MV allocation by explicitly incorporating an uncertainty model on the input data or parameters in a portfolio selection problem and carrying out the analysis for the worst-case scenario under this model. Recent work on robust portfolio optimization includes [9, 10, 11, 12, 18].

In this section, we consider the robust counterpart of the SR maximization problem. The reader is referred to [16] for the importance of this problem in robust MV analysis. In this paper, we focus on the computational aspects of the robust counterpart.

We assume that the expected return μ and covariance Σ of the asset returns are uncertain but known to belong to a convex compact subset $\mathcal{U}$ of $\mathbb{R}^n \times \mathbb{S}^n_{++}$. We also assume there exists an admissible portfolio $\bar{w} \in \mathcal{W}$ of risky assets whose worst-case mean return is greater than the risk-free return:

$$\text{there exists a portfolio } \bar{w} \in \mathcal{W} \text{ such that } \mu^T w > \mu_{\mathrm{rf}} \text{ for all } (\mu, \Sigma) \in \mathcal{U}. \tag{23}$$

Worst-case SR maximization. The zero-sum game of choosing w from $\mathcal{W}$, to maximize the SR, and choosing (μ, Σ) from $\mathcal{U}$, to minimize the SR, is associated with the following two problems:

- worst-case SR maximization problem of finding an admissible portfolio w that maximizes the worst-case SR (over the given model $\mathcal{U}$ of uncertainty)

$$\begin{array}{ll} \text{maximize} & \inf_{(\mu,\Sigma)\in\mathcal{U}} S(w,\mu,\Sigma) \\ \text{subject to} & w \in \mathcal{W}, \end{array} \tag{24}$$

- worst-case market price of risk analysis (MPRA) problem of finding the least-favorable statistics (over the uncertainty set $\mathcal{U}$), with portfolio weights chosen optimally for the asset return statistics,

$$\begin{array}{ll} \text{minimize} & \sup_{w\in\mathcal{W}} S(w,\mu,\Sigma) \\ \text{subject to} & (\mu,\Sigma) \in \mathcal{U}. \end{array} \tag{25}$$

The SR is not a fractional function of the form (1), so we cannot apply Theorem 1 directly to the zero-sum game given above. We can get around this difficulty by using the fact that when the domain is restricted to $\mathcal{W}$, the SR has the form (1)

$$\frac{\mu^T w - \mu_{\mathrm{rf}}}{\sqrt{w^T \Sigma w}} = \frac{w^T(\mu - \mu_{\mathrm{rf}}\mathbf{1})}{\sqrt{w^T \Sigma w}} = f(w, \mu - \mu_{\mathrm{rf}}\mathbf{1}, \Sigma) \qquad \forall w \in \mathcal{W}$$

and

$$S(w,\mu,\Sigma) = f(tw, \mu - \mu_{\mathrm{rf}}\mathbf{1}, \Sigma), \quad w \in \mathcal{W}, \quad t > 0, \tag{26}$$

whenever $S(w,\mu,\Sigma) > 0$.

The set

$$\mathcal{X} = \mathrm{cl}\,\{tw \in \mathbb{R}^n \mid w \in \mathcal{W},\ t > 0\} \backslash \{0\},$$

where $\operatorname{cl} A$ means the closure of the set A and $A\backslash B$ means the complement of B in A, is a cone in $\mathbb{R}^n$, with $\mathcal{X} \cup \{0\}$ closed and convex. Assumption (23), along with the compactness of $\mathcal{U}$, means that

$$\inf_{(\mu,\Sigma)\in\mathcal{U}} \bar{w}^T(\mu - \mu_{\rm rf}\mathbf{1}) > 0.$$

We can therefore apply Theorem 1 to the zero-sum game of choosing w from $\mathcal{X}$, to maximize $f(x, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma)$, and choosing (μ, Σ) from $\mathcal{U}$, to minimize $f(x, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma)$.

The max-min and min-max problems associated with the game are

- max-min problem

$$\begin{array}{ll} \text{maximize} & \inf_{(\mu,\Sigma)\in\mathcal{U}} f(x, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma) \\ \text{subject to} & x \in \mathcal{X}, \end{array} \tag{27}$$

- min-max problem

$$\begin{array}{ll} \text{minimize} & \sup_{x\in\mathcal{X}} f(x, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma) \\ \text{subject to} & (\mu, \Sigma) \in \mathcal{U}. \end{array} \tag{28}$$

According to Theorem 1, the two problems have the same optimal value:

$$\sup_{x\in\mathcal{X}} \inf_{(\mu,\Sigma)\in\mathcal{U}} f(x, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma) = \inf_{(\mu,\Sigma)\in\mathcal{U}} \sup_{x\in\mathcal{X}} f(x, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma). \tag{29}$$

As a result, the SR satisfies the minimax equality

$$\sup_{w\in\mathcal{W}} \inf_{(\mu,\Sigma)\in\mathcal{U}} S(w, \mu, \Sigma) = \inf_{(\mu,\Sigma)\in\mathcal{U}} \sup_{w\in\mathcal{W}} S(w, \mu, \Sigma),$$

which follows from (26) and (29).

From Proposition 1, we can see that the min-max problem (28) is equivalent to the convex problem

$$\begin{array}{ll} \text{minimize} & (\mu - \mu_{\rm rf}\mathbf{1} + \lambda)^T \Sigma^{-1} (\mu - \mu_{\rm rf}\mathbf{1} + \lambda) \\ \text{subject to} & (\mu, \Sigma) \in \mathcal{U}, \quad \lambda \in \mathcal{W}^{\oplus} \end{array} \tag{30}$$

in which the optimization variables are $\mu \in \mathbb{R}^n$, $\Sigma = \Sigma^T \in \mathbb{R}^{n\times n}$, and $\lambda \in \mathbb{R}^n$. Here $\mathcal{W}^{\oplus}$ is the positive conjugate cone $\mathcal{W}$, which is equal to the dual cone $X^\star$ of $\mathcal{X}$:

$$\mathcal{X}^* = \mathcal{W}^{\oplus} = \left\{ \lambda \in \mathbb{R}^n \mid \lambda^T w \geq 0 \ \forall w \in \mathcal{W} \right\}.$$

The convex problem (30) has a solution, say, $(\mu^\star, \Sigma^\star, \lambda^\star)$. Then,

$$x^\star = {\Sigma^\star}^{-1}(\mu^\star - \mu_{\rm rf}\mathbf{1} + \lambda^\star) \in \mathcal{X}$$

is a unique solution of the max-min problem (27) (up to positive scaling). Moreover, the saddle-point property

$$f(x, \mu^\star - \mu_{\rm rf}\mathbf{1}, \Sigma^\star) \leq f(x^\star, \mu^\star - \mu_{\rm rf}\mathbf{1}, \Sigma^\star) \leq f(x^\star, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma), \quad x \in \mathcal{X}, \quad (\mu, \Sigma) \in \mathcal{U}, \tag{31}$$

holds. We can see from (26) that

$$S(w, \mu^\star, \Sigma^\star) \leq S(w^\star, \mu^\star, \Sigma^\star) \leq S(w^\star, \mu, \Sigma) \quad \forall w \in \mathcal{W} \quad \forall (\mu, \Sigma) \in \mathcal{U}. \tag{32}$$

Finally, since $\mathbf{1}^T x \geq 0$ for all $x \in \mathcal{X}$, we have

$$\mathbf{1}^T \Sigma^{\star -1}(\mu^\star - \mu_{\mathrm{rf}}\mathbf{1} + \lambda^\star) \geq 0.$$

If $x^\star$ satisfies $\mathbf{1}^T x^\star > 0$, the portfolio

$$w^\star = \left(1/\mathbf{1}^T x^\star\right) x^\star = \frac{1}{\mathbf{1}^T \Sigma^{\star -1}(\mu^\star - \mu_{\mathrm{rf}}\mathbf{1} + \lambda^\star)} \Sigma^{\star -1}(\mu^\star - \mu_{\mathrm{rf}}\mathbf{1} + \lambda^\star) \tag{33}$$

satisfies the budget constraint and is admissible (i.e., $w^\star \in \mathcal{W}$); i.e., it is a solution to the worst-case SR maximization (24). Moreover, it is the unique solution to the worst-case SR maximization (24). The case of $\mathbf{1}^T x^\star = 0$ may arise when the set $\mathcal{W}$ is unbounded. In this case, the worst-case SR maximization problem (24) has no solution, so the game involving the SR has no saddle point.

Minimax properties of the SR. The results established above are summarized in the following proposition.

PROPOSITION 3. *Suppose that the uncertainty set $\mathcal{U}$ is compact and convex. Suppose further that Assumption* (23) *holds. Let $(\mu^\star, \Sigma^\star, \lambda^\star)$ be a solution to the convex problem* (30). *Then, we have the following:*

(i) *If $\mathbf{1}^T \Sigma^{-1}(\mu - \mu_{\mathrm{rf}}\mathbf{1} + \lambda^\star) > 0$, then the triple $(w^\star, \mu^\star, \Sigma^\star)$ with $w^\star$ in* (33) *satisfies the saddle-point property* (32), *and $w^\star$ is the unique solution to the worst-case SR maximization problem* (24).

(ii) *If $\mathbf{1}^T \Sigma^{-1}(\mu - \mu_{\mathrm{rf}}\mathbf{1} + \lambda^\star) = 0$, then the optimal value of the worst-case SR maximization problem* (24) *is not achieved by any portfolio in $\mathcal{W}$.*

Moreover, the minimax equality

$$\sup_{w \in \mathcal{W}} \inf_{(\mu,\Sigma) \in \mathcal{U}} S(w, \mu, \Sigma) = \inf_{(\mu,\Sigma) \in \mathcal{U}} \sup_{w \in \mathcal{W}} S(w, \mu, \Sigma)$$

holds regardless of the existence of a solution.

The worst-case MPRA problem (25) is equivalent to the min-max problem (28), which is in turn equivalent to the convex problem (30). This proposition shows that the TP of the least-favorable model $(\mu^\star, \Sigma^\star)$ solves the worst-case SR maximization problem (24). The saddle-point property (32) means that the portfolio $w^\star$ in (33) is the TP of the least-favorable model $(\mu^\star, \Sigma^\star)$. The portfolio is called the *robust TP*.

5.3. Numerical example. We illustrate the result with a synthetic example, with $n = 7$ risky assets. The risk-free return is taken as $\mu_{\mathrm{rf}} = 5$.

Setup. The nominal returns $\bar{\mu}_i$ and variances $\bar{\sigma}_i^2$ of the risky assets are taken as

$$\bar{\mu} = [10.3 \;\; 10.5 \;\; 5.5 \;\; 10.5 \;\; 110 \;\; 14.4 \;\; 10.1]^T,$$
$$\bar{\sigma} = [11.3 \;\; 18.1 \;\; 6.8 \;\; 22.7 \;\; 24.0 \;\; 14.7 \;\; 20.9]^T.$$

The nominal correlation matrix $\bar{\Omega}$ is

$$\bar{\Omega} = \begin{bmatrix} 1.00 & .07 & -.12 & .43 & -.11 & .44 & .25 \\ & 1.00 & .73 & -.14 & .39 & .28 & .10 \\ & & 1.00 & .14 & .50 & .52 & -.13 \\ & & & 1.00 & .04 & .35 & .38 \\ & & & & 1.00 & .70 & .04 \\ & & & & & 1.00 & -.09 \\ & & & & & & 1.00 \end{bmatrix}.$$

The nominal covariance is

$$\bar{\Sigma} = \operatorname{diag}(\bar{\sigma})\bar{\Omega}\operatorname{diag}(\bar{\sigma}),$$

where we use $\operatorname{diag}(u_1, \ldots, u_m)$ to denote the diagonal matrix with diagonal entries $u_1, \ldots, u_m$.

The mean uncertainty model used in our study is

$$\begin{aligned} |\mu_i - \bar{\mu}_i| &\le 0.3|\bar{\mu}_i|, \quad i = 1, \ldots, 7, \\ \left|\mathbf{1}^T\mu - \mathbf{1}^T\bar{\mu}\right| &\le 0.15\left|\mathbf{1}^T\bar{\mu}\right|. \end{aligned}$$

These constraints mean that the possible variation in the expected return of each asset is at most 30%, and the possible variation in the expected return of the portfolio $(1/n)\mathbf{1}$ (in which a fraction $1/n$ of the budget is allocated to each asset of the n assets) is at most 15%. The covariance uncertainty model used in our study is

$$\begin{aligned} \left|\Sigma_{ij} - \bar{\Sigma}_{ij}\right| &\le 0.3\left|\bar{\Sigma}_{ij}\right|, \quad i, j = 1, \ldots, 7, \\ \left\|\Sigma - \bar{\Sigma}\right\|_F &\le 0.15\left\|\bar{\Sigma}\right\|_F. \end{aligned}$$

(Here, $\|A\|_F$ denotes the Frobenius norm of A, i.e., $\|A\|_F = (\sum_{i,j=1}^n A_{ij}^2)^{1/2}$.) These constraints mean that the possible variation in each component of the covariance matrix is at most 30% and the possible deviation of the covariance from the nominal covariance is at most 15% in terms of the Frobenius norm.

We consider the case when short selling is allowed in a limited way as follows:

$$w = w_{\text{long}} - w_{\text{short}}, \quad w_{\text{long}}, w_{\text{short}} \succeq 0, \quad \mathbf{1}^T w_{\text{short}} \le \eta \mathbf{1}^T w_{\text{long}}, \tag{34}$$

where η is a positive constant and w_{long} and w_{short} represent the total long and short positions at the beginning of the period, respectively. ($w \succeq 0$ means that w is componentwise nonnegative.) The last constraint limits the total short position to some fraction η of the total long position. In our numerical study, we take $\gamma = 0.3$.

The asset constraint set is given by the cone

$$\mathcal{W} = \left\{ w \in \mathbb{R}^n \mid w = w_{\text{long}} - w_{\text{short}},\ A \begin{bmatrix} w_{\text{long}} \\ w_{\text{short}} \end{bmatrix} \preceq 0 \right\},$$

where

$$A = \begin{bmatrix} -I & 0 \\ 0 & -I \\ -\gamma\mathbf{1}^T & \mathbf{1}^T \end{bmatrix} \in \mathbb{R}^{(2n+1)\times(2n)}.$$

A simple argument based on linear programming duality shows that the dual cone of $\mathcal{X} = \mathcal{W}$ is given by

$$\mathcal{X}^* = \left\{ \lambda \in \mathbb{R}^n \;\middle|\; \text{there exists } y \succeq 0 \text{ such that } A^T y + \begin{bmatrix} \lambda \\ -\lambda \end{bmatrix} = 0 \right\}.$$

Comparison results. We can find the robust TP by applying Theorem 1 to the corresponding problem (27) with the asset allocation constraints and uncertainty model described above. The nominal TP can be found using Theorem 1 with the singleton $\mathcal{U} = \{(\bar{\mu}, \bar{\Sigma})\}$.

Table 3
Nominal and worst-case SRs of the nominal and robust TPs.

	Nominal SR	Worst-case SR
Nominal TP	0.74	0.22
Robust TP	0.57	0.36

Table 4
Outperformance probability of the nominal and robust TPs.

	$\mathbf{P}_{\rm nom}$	$\mathbf{P}_{\rm wc}$
Nominal TP	0.77	0.59
Robust TP	0.71	0.64

Table 3 shows the nominal and worst-case SRs of the nominal optimal and robust optimal allocations. In comparison with the market portfolio, the robust market portfolio shows a relatively small decrease in the SR, in the presence of possible variations in the parameters. The SR of the robust market portfolio decreases about 39% from 0.57 to 0.36, while the SR of the nominal market portfolio decreases about 70% from 0.74 to 0.22.

Table 4 shows the probabilities of outperforming the risk-free asset for the nominal optimal and robust optimal weight allocations, when the asset returns follow a normal distribution. Here, $\mathbf{P}_{\rm nom}$ is the probability of beating the risk-free asset without uncertainty, called the outperformance probability, and $\mathbf{P}_{\rm wc}$ is the worst-case probability of outperforming the risk-free asset with uncertainty. The nominal optimal TP achieves $\mathbf{P}_{\rm nom} = 0.77$, which corresponds to 77% of outperforming the risk-free asset without uncertainty. However, in the presence of uncertainty in the parameters, its performance degrades rapidly; the worst-case outperformance probability for the nominal optimal discriminant is 59%. The robust optimal allocation performs well in the presence of uncertainty in the parameters, with the worst-case outperformance probability 5% higher than that of the nominal optimal allocation.

6. Conclusions. The fractional function $f(x, a, B) = a^T x/\sqrt{x^T Bx}$ comes up in many contexts, some of which are discussed above. In this paper, we have established a minimax result for this function and a general computational method, based on convex optimization, for computing a saddle point.

The arguments used to establish the minimax result do not appear to be extensible to other fractional functions that have a similar form. For instance, the extension to a general fractional function of the form

$$g(x, A, B) = \frac{x^T Ax}{x^T Bx},$$

which is the Rayleigh quotient of the matrix pair $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ evaluated at $x \in \mathbb{R}^n$, is not possible; see, e.g., [31] for a counterexample. However, the arguments can be extended to the special case when A is a dyad, i.e., $A = aa^T$, with $a \in \mathbb{R}^n$, and $\mathcal{X} = \mathbb{R}^n\backslash\{0\}$. In this case, the minimax equality

$$\sup_{x\neq 0} \inf_{(a,B)\in\mathcal{U}} \frac{\left(x^T a\right)^2}{x^T Bx} = \inf_{(a,B)\in\mathcal{U}} \sup_{x\neq 0} \frac{\left(x^T a\right)^2}{x^T Bx}$$

holds with assumption (7); see [17] for the proof.

Appendix A. Proofs.

A.1. Proof of Proposition 1. We first show that the optimal value of (8) is positive. We start by noting that

$$\inf_{(a,B)\in\mathcal{U}} \frac{\bar{x}^T a}{\sqrt{\bar{x}^T B \bar{x}}} > 0, \tag{35}$$

with $\bar{x}$ in (7), and

$$\inf_{(a,B)\in\mathcal{U},\lambda\in\mathcal{X}^*} \frac{\bar{x}^T(a+\lambda)}{\sqrt{\bar{x}^T B\bar{x}}} = \inf_{(a,B)\in\mathcal{U}} \inf_{\lambda\in\mathcal{X}^*} \frac{\bar{x}^T(a+\lambda)}{\sqrt{\bar{x}^T B\bar{x}}} = \inf_{(a,B)\in\mathcal{U}} \frac{\bar{x}^T a}{\sqrt{\bar{x}^T B\bar{x}}}. \tag{36}$$

Here, we have used (35) and $\inf_{\lambda\in\mathcal{X}^*} \bar{x}^T\lambda = 0$. By the Cauchy–Schwarz inequality, $x^T(a+\lambda)/\sqrt{x^T Bx}$ is maximized over nonzero x by $x = B^{-1}(a+\lambda)$, so

$$\sup_{x\neq 0} x^T(a+\lambda)/\sqrt{x^T Bx} = \left[(a+\lambda)^T B^{-1}(a+\lambda)\right]^{1/2}.$$

It follows from the minimax inequality (5), (35), and (36) that

$$\begin{aligned}
\inf_{(a,B)\in\mathcal{U},\lambda\in\mathcal{X}^*} \left[(a+\lambda)^T B^{-1}(a+\lambda)\right]^{1/2} &= \inf_{(a,B)\in\mathcal{U},\lambda\in\mathcal{X}^*} \sup_{x\neq 0} \frac{x^T(a+\lambda)}{\sqrt{x^T Bx}} \\
&\geq \sup_{x\neq 0} \inf_{(a,B)\in\mathcal{U},\lambda\in\mathcal{X}^*} \frac{x^T(a+\lambda)}{\sqrt{x^T Bx}} \\
&\geq \inf_{(a,B)\in\mathcal{U},\lambda\in\mathcal{X}^*} \frac{\bar{x}^T(a+\lambda)}{\sqrt{\bar{x}^T B\bar{x}}} \\
&> 0.
\end{aligned}$$

(Here, we use the fact that the weak minimax property for $x^T(a+\lambda)/\sqrt{x^T Bx}$ holds for any $\mathcal{U} \subseteq \mathbb{R}^n \times \mathbb{S}^n_{++}$ and $\mathcal{X} \subseteq \mathbb{R}^n$.)

We next show that (8) has a solution. There is a sequence

$$\left\{ \left(a^{(i)} + \lambda^{(i)}, B^{(i)}\right) \;\middle|\; \left(a^{(i)}, B^{(i)}\right) \in \mathcal{U},\ \lambda^{(i)} \in \mathcal{X}^*,\ i = 1, 2, \ldots \right\}$$

such that

$$\lim_{i\to\infty} \left(a^{(i)} + \lambda^{(i)}\right)^T {B^{(i)}}^{-1} \left(a^{(i)} + \lambda^{(i)}\right) = \inf_{(a,B)\in\mathcal{U},\lambda\in\mathcal{X}^*} (a+\lambda)^T B^{-1}(a+\lambda). \tag{37}$$

Since $\mathcal{U}$ is a compact subset of $\mathbb{R}^n \times \mathbb{S}^n_{++}$, we have

$$\sup\left\{\lambda_{\max}\left(B^{-1}\right) \;\middle|\; \text{for all } B \text{ with } (a,B)\in\mathcal{U}\right\} < \infty.$$

(Here $\lambda_{\max}(B)$ is the maximum eigenvalue of B.) Then, $S_1 = \{a^{(i)} + \lambda^{(i)} \in \mathbb{R}^n \mid i = 1, 2, \ldots\}$ must be bounded. (Otherwise, there arises a contradiction to (37).) Since $\mathcal{U}$ is compact, the sequence $S_2 = \{(a^{(i)}, B^{(i)}) \in \mathcal{U} \mid i = 1, 2, \ldots\}$ is bounded, which along with the boundedness of S_1 means that $S_3 = \{\lambda^{(i)} \in \mathbb{R}^n \mid i = 1, 2, \ldots\}$ is also bounded. The bounded sequences S_2 and S_3 have convergent subsequences, which converge to, say, $(a^\star, B^\star)$ and $\lambda^\star$, respectively. Since $\mathcal{U}$ and $\mathcal{X}^*$ are closed, $(a^\star, B^\star) \in \mathcal{U}$ and $\lambda^\star \in \mathcal{X}^*$. The triple $(a^\star, B^\star, \lambda^\star)$ achieves the optimal value of (8). Since the optimal value is positive, $a^\star + \lambda^\star \neq 0$.

The equivalence between (4) and (8) follows from the following implication:

$$\sup_{x\in\mathcal{X}} x^T a > 0 \qquad \Longrightarrow \qquad \sup_{x\in\mathcal{X}} \frac{x^T a}{\sqrt{x^T B x}} = \inf_{\lambda\in\mathcal{X}^*} \left[(a+\lambda)^T B^{-1}(a+\lambda)\right]^{1/2}. \tag{38}$$

Then, (4) is equivalent to

$$\begin{array}{ll} \text{minimize} & \inf_{\lambda\in\mathcal{X}^*} (a+\lambda)^T B^{-1}(a+\lambda) \\ \text{subject to} & (a,B)\in\mathcal{U}. \end{array}$$

It is now easy to see that (4) is equivalent to (8).

To establish the implication, we show that

$$\sup_{x\in\mathcal{X}} \frac{x^T a}{\sqrt{x^T B x}} = \sup_{x\neq 0} \inf_{\lambda\in\mathcal{X}^*} \frac{x^T(a+\lambda)}{\sqrt{x^T B x}}. \tag{39}$$

First, suppose that $x \in \mathcal{X}$. Then, $\lambda^T x \geq 0$ for any $\lambda \in \mathcal{X}^*$ and $0 \in \mathcal{X}^*$, so $\inf_{\lambda\in\mathcal{X}^*} \lambda^T x = 0$. Thus,

$$\inf_{\lambda\in\mathcal{X}^*} \frac{x^T(a+\lambda)}{\sqrt{x^T B x}} = \frac{x^T a}{\sqrt{x^T B x}}.$$

Next, suppose that $x \notin \mathcal{X}\cup\{0\}$. Note from $\mathcal{X}^{**} = \mathcal{X}\cup\{0\}$ that there exists a nonzero $\bar\lambda \in \mathcal{X}^*$, with $\bar\lambda^T x < 0$. Then,

$$\inf_{\lambda\in\mathcal{X}^*} \frac{x^T(a+\lambda)}{\sqrt{x^T B x}} \leq \inf_{t>0} \left(\frac{x^T a}{\sqrt{x^T B x}} + \frac{t x^T \bar\lambda}{\sqrt{x^T B x}} \right) = -\infty \quad \forall x \notin \mathcal{X}\cup\{0\}.$$

When $\sup_{x\in\mathcal{X}} x^T a > 0$, we have from (39) that

$$\inf_{\lambda\in\mathcal{X}^*} \sup_{x\neq 0} \frac{x^T(a+\lambda)}{\sqrt{x^T B x}} > 0.$$

By the Cauchy–Schwarz inequality,

$$\inf_{\lambda\in\mathcal{X}^*} \sup_{x\neq 0} \frac{x^T(a+\lambda)}{\sqrt{x^T B x}} = \inf_{\lambda\in\mathcal{X}^*} \left[(a+\lambda)^T B^{-1}(a+\lambda)\right]^{1/2} = \left[\inf_{\lambda\in\mathcal{X}^*} (a+\lambda)^T B^{-1}(a+\lambda)\right]^{1/2}.$$

Since $(a+\lambda)^T B^{-1}(a+\lambda)$ is strictly concave in λ, we can see that there is $\lambda^\star$ such that

$$\inf_{\lambda\in\mathcal{X}^*} (a+\lambda)^T B^{-1}(a+\lambda) = (a+\lambda^\star)^T B^{-1}(a+\lambda^\star). \tag{40}$$

Then,

$$\sup_{x\neq 0} \frac{x^T(a+\lambda^\star)}{\sqrt{x^T B x}} = \inf_{\lambda\in\mathcal{X}^*} \sup_{x\neq 0} \frac{x^T(a+\lambda)}{\sqrt{x^T B x}}.$$

As will be seen soon, $x^\star = B^{-1}(a+\lambda^\star)$ satisfies

$$\frac{x^{\star T}(a+\lambda^\star)}{\sqrt{x^{\star T} B x^\star}} = \inf_{\lambda\in\mathcal{X}^*} \frac{x^{\star T}(a+\lambda)}{\sqrt{x^{\star T} B x^\star}}. \tag{41}$$

Therefore,

$$\sup_{x\neq 0}\inf_{\lambda\in\mathcal{X}^*}\frac{x^T(a+\lambda)}{\sqrt{x^TBx}}=\inf_{\lambda\in\mathcal{X}^*}\sup_{x\neq 0}\frac{x^T(a+\lambda)}{\sqrt{x^TBx}}.$$

Taken together, the results established above show that

$$\begin{aligned}\sup_{x\in\mathcal{X}}\frac{x^Ta}{\sqrt{x^TBx}}&=\sup_{x\neq 0}\inf_{\lambda\in\mathcal{X}^*}\frac{x^T(a+\lambda)}{\sqrt{x^TBx}}=\inf_{\lambda\in\mathcal{X}^*}\sup_{x\neq 0}\frac{x^T(a+\lambda)}{\sqrt{x^TBx}}\\&=\inf_{\lambda\in\mathcal{X}^*}\left[(a+\lambda)^TB^{-1}(a+\lambda)\right]^{1/2}.\end{aligned}$$

We complete the proof by establishing (41). To this end, we derive explicitly the optimality condition for $\lambda^\star$ to satisfy (40):

$$2x^{\star T}(\lambda-\lambda^\star)\geq 0\quad\forall\lambda\in\mathcal{X}^*,\tag{42}$$

with $x^\star=B^{\star-1}(a+\lambda^\star)$. (See [7, section 4.2.3].) We now show that $x^\star$ satisfies (41). To this end, we note that $\bar{\lambda}$ is optimal for (41) if and only if

$$\left\langle \nabla_\lambda\frac{\left(x^{\star T}(a+\lambda)\right)^2}{x^{\star T}Bx^\star}\bigg|_{\bar{\lambda}},\left(\lambda-\bar{\lambda}\right)\right\rangle\geq 0\quad\forall\lambda\in\mathcal{X}^*.$$

Here $\nabla_\lambda h(\lambda)|_{\bar{\lambda}}$ denotes the gradient of h at the point $\bar{\lambda}$. We can write the optimality condition as

$$2\frac{x^{\star T}\left(a+\bar{\lambda}\right)}{x^{\star T}\bar{B}x^\star}x^{\star T}\left(\lambda-\bar{\lambda}\right)\geq 0\quad\forall\lambda\in\mathcal{X}^*.$$

Substituting $\bar{\lambda}=\lambda^\star$ and noting that $(a+\lambda)^{\star T}x^\star/x^{\star T}B^\star x^\star=1$, the optimality condition reduces to (42). Thus, we have shown that $\lambda^\star$ is optimal for (41).

A.2. Proof of Theorem 1. We will establish the following claims:

- $x^\star=B^{\star-1}(a^\star+\lambda^\star)\in\mathcal{X}$.
- $(x^\star,a^\star,\lambda^\star,B^\star)$ satisfies the saddle-point property

$$\frac{x^T(a^\star+\lambda^\star)}{\sqrt{x^TB^\star x}}\leq\frac{x^{\star T}(a^\star+\lambda^\star)}{\sqrt{x^{\star T}B^\star x^\star}}\leq\frac{x^{\star T}(a+\lambda)}{\sqrt{x^{\star T}Bx^\star}}\quad\forall x\neq 0\;\;\forall\lambda\in\mathcal{X}^*\;\;\forall(a,B)\in\mathcal{U}.\tag{43}$$

- $x^\star$ and $\lambda^\star$ are orthogonal to each other:

$$x^{\star T}\lambda^\star=0.\tag{44}$$

The claims of Theorem 1 follow directly from the claims above. By definition of the dual cone, we have $\lambda^{\star T}x\geq 0$ for all $x\in\mathcal{X}$ and $0\in\mathcal{X}^*$. It follows from (43) and (44) that

$$\begin{aligned}\frac{x^Ta^\star}{\sqrt{x^TB^\star x}}&\leq\frac{x^T(a^\star+\lambda^\star)}{\sqrt{x^TB^\star x}}\leq\frac{x^{\star T}(a^\star+\lambda^\star)}{\sqrt{x^{\star T}B^\star x^\star}}=\frac{x^{\star T}a^\star}{\sqrt{x^{\star T}B^\star x^\star}}\\&\leq\frac{x^{\star T}a}{\sqrt{x^{\star T}Bx^\star}}\;\forall x\in\mathcal{X}\;\forall(a,B)\in\mathcal{U}.\end{aligned}$$

The saddle-point property (43) is equivalent to showing that

$$\sup_{x\neq 0} \frac{x^T(a^\star+\lambda^\star)}{\sqrt{x^T B^\star x}} = \frac{x^{\star T}(a^\star+\lambda^\star)}{\sqrt{x^{\star T} B^\star x^\star}} \tag{45}$$

and

$$\inf_{(a,B)\in\mathcal{U},\lambda\in\mathcal{X}^*} \frac{x^{\star T}(a+\lambda)}{\sqrt{x^{\star T} B x^\star}} = \frac{x^{\star T}(a^\star+\lambda^\star)}{\sqrt{x^{\star T} B^\star x^\star}}. \tag{46}$$

Here, (45) follows from the Cauchy–Schwarz inequality.

We establish (46) by showing an equivalent claim

$$\inf_{(c,B)\in\mathcal{V}} \frac{x^{\star T} c}{\sqrt{x^{\star T} B x^\star}} = \frac{x^{\star T} c^\star}{\sqrt{x^{\star T} B^\star x^\star}},$$

where

$$c^\star = a^\star + \lambda^\star, \qquad \mathcal{V} = \{(a+\lambda, B) \in \mathbb{R}^n \times \mathbb{S}^n_{++} \mid (a,B)\in\mathcal{U},\ \lambda\in\mathcal{X}^*\}.$$

The set $\mathcal{V}$ is closed and convex.

We know that $(c^\star, B^\star)$ is optimal for the convex problem

$$\begin{array}{ll} \text{minimize} & g(c,B) = c^T B^{-1} c \\ \text{subject to} & (c,B)\in\mathcal{V}, \end{array} \tag{47}$$

with variables $c \in \mathbb{R}^n$ and $B = B^T \in \mathbb{R}^{n\times n}$. From the optimality condition of this problem that $(c^\star, B^\star)$ satisfies, we will prove that $(c^\star, B^\star)$ is also optimal for the problem

$$\begin{array}{ll} \text{minimize} & x^{\star T} c/\sqrt{x^{\star T} B x^\star} \\ \text{subject to} & (c,B)\in\mathcal{V}, \end{array} \tag{48}$$

with variables $c \in \mathbb{R}^n$ and $B = B^T \in \mathbb{R}^{n\times n}$. The proof is based on an extension of the arguments used to establish (41).

We derive explicitly the optimality condition for the convex problem (47). The pair $(c^\star, B^\star)$ must satisfy the optimality condition

$$\left\langle \nabla_c g(c,B)|_{(c^\star,B^\star)}, (c-c^\star)\right\rangle + \left\langle \nabla_B g(c,B)|_{(c^\star,B^\star)}, (B-B^\star)\right\rangle \geq 0 \quad \forall (c,B)\in\mathcal{V}$$

(see [7, section 4.2.3]). Here $(\nabla_c f(c,B)|_{(\bar{c},\bar{B})}, \nabla_B g(c,B)|_{(\bar{c},\bar{B})})$ denotes the gradient of f at the point (c,B). Using $\nabla_c(c^T B^{-1} c) = 2B^{-1}c$, $\nabla_B(c^T B^{-1} c) = -B^{-1}cc^T B^{-1}$, and $\langle X, Y\rangle = \mathbf{Tr}(XY)$ for $X, Y \in \mathbb{S}^n$, where $\mathbf{Tr}$ denotes trace, we can express the optimality condition as

$$2c^{\star T} B^{\star -1}(c-c^\star) - \mathbf{Tr}\, B^{\star -1} c^\star c^{\star T} B^{\star -1}(B - B^\star) \geq 0 \quad \forall (c,B)\in\mathcal{V}$$

or equivalently

$$2x^{\star T}(c-c^\star) - x^{\star T}(B-B^\star)x^\star \geq 0 \quad \forall (c,B)\in\mathcal{V}, \tag{49}$$

with $x^\star = B^{\star -1} c^\star$.

To establish the optimality of $(c^\star, B^\star)$ for (48), we show that a solution of (48) is also a solution to the optimization problem

$$\begin{array}{ll} \text{minimize} & \left(x^{\star T} c\right)^2 / \left(x^{\star T} B x^\star\right) \\ \text{subject to} & (c, B) \in \mathcal{V}, \end{array} \tag{50}$$

with variables $c \in \mathbb{R}^n$ and $B = B^T \in \mathbb{R}^{n \times n}$ and vice versa. To show that (50) is a convex optimization problem, we must show that the objective is a convex function of c and B. To do so, we express the objective as the composition

$$\frac{\left(x^{\star T} c\right)^2}{x^{\star T} B x^\star} = g(H(c, B)),$$

where $g(u, t) = u^2/t$ and H is the function

$$H(c, B) = \left(x^{\star T} c, x^{\star T} B x^\star\right).$$

The function H is linear (as a mapping from c and B into $\mathbb{R}^2$), and the function g is convex (provided $t > 0$, which holds here). Thus, the composition f is a convex function of a and B. (See [7, section 3].)

This equivalence between (48) and (50) follows from

$$x^{\star T} c / \left(x^{\star T} B x^\star\right)^{1/2} > 0 \quad \forall (c, B) \in \mathcal{V},$$

which is a direct consequence of the optimality condition (49):

$$\begin{aligned} 2x^{\star T} c &\geq 2x^{\star T} c^\star + x^{\star T}(B - B^\star)x^\star \\ &= x^{\star T} c^\star + x^{\star T}(c^\star - B^\star x^\star) + x^{\star T} B x^\star \\ &= x^{\star T} B^{\star -1} x^\star + x^{\star T} B x^\star \\ &> 0 \quad \forall (c, B) \in \mathcal{V}. \end{aligned}$$

We now show that $(c^\star, B^\star)$ is optimal for (50) and hence for (48). The optimality condition for (50) is that a pair $(\bar{c}, \bar{B})$ is optimal for (50) if and only if

$$\left\langle \nabla_c \frac{\left(x^{\star T} c\right)^2}{x^{\star T} B x^\star}\bigg|_{(\bar{c}, \bar{B})}, (c - \bar{c}) \right\rangle + \left\langle \nabla_B \frac{\left(x^{\star T} c\right)^2}{x^{\star T} B x^\star}\bigg|_{(\bar{c}, \bar{B})}, \left(B - \bar{B}\right) \right\rangle \geq 0 \quad \forall (c, B) \in \mathcal{V}$$

(see [7, section 4.2.3]). Using

$$\nabla_c \frac{\left(x^{\star T} c\right)^2}{x^{\star T} B x^\star} = 2\frac{c^T x^\star}{x^{\star T} B x^\star} x^\star, \qquad \nabla_B \frac{\left(x^{\star T} c\right)^2}{x^{\star T} B x^\star} = -\frac{\left(c^T x^\star\right)^2}{\left(x^{\star T} B x^\star\right)^2} x^\star x^{\star T},$$

we can write the optimality condition as

$$\begin{aligned} & 2\frac{x^{\star T} \bar{c}}{x^{\star T} \bar{B} x^\star} x^{\star T}(c - \bar{c}) - \mathbf{Tr}\, \frac{\left(x^{\star T} \bar{c}\right)^2}{\left(x^{\star T} \bar{B} x^\star\right)^2} x^\star x^{\star T} \left(B - \bar{B}\right) \\ & = 2\frac{x^{\star T} \bar{c}}{x^{\star T} \bar{B} x^\star} x^{\star T}(c - \bar{c}) - \frac{\left(x^{\star T} \bar{c}\right)^2}{\left(x^{\star T} \bar{B} x^\star\right)^2} x^{\star T} \left(B - \bar{B}\right) x^\star \\ & \geq 0 \quad \forall (c, B) \in \mathcal{V}. \end{aligned}$$

Substituting $\bar{c} = c^\star$, $\bar{B} = B^\star$, and noting that $c^{\star T}x^\star / x^{\star T}B^\star x^\star = 1$, the optimality condition reduces to

$$2x^{\star T}(c - c^\star) - x^{\star T}(B - B^\star)x^\star \geq 0 \quad \forall (c, B) \in \mathcal{V},$$

which is precisely (49). Thus, we have shown that $(c^\star, B^\star)$ is optimal for (50), which in turn means that it is also optimal for (48).

We next show by way of contradiction that $x^\star \in \mathcal{X}$. Suppose that $x^\star \notin \mathcal{X}$. Then, it follows from $\mathcal{X}^{**} = \mathcal{X} \cup \{0\}$ that there is $\bar{\lambda} \in \mathcal{X}^*$ such that $\bar{\lambda}^T x^\star < 0$. For any fixed $(\bar{a}, \bar{B})$ in $\mathcal{U}$, we can see from (43) (already established) that

$$\inf_{(a,B)\in\mathcal{U},\, \lambda\in\mathcal{X}^*} \frac{x^{\star T}(a+\lambda)}{\sqrt{x^{\star T}Bx^\star}} \leq \inf_{\lambda\in\mathcal{X}^*} \frac{x^{\star T}(\bar{a}+\lambda)}{\sqrt{x^{\star T}\bar{B}x^\star}} \leq \frac{x^{\star T}\bar{a}}{\sqrt{x^{\star T}\bar{B}x^\star}} + \inf_{t\geq 0} \frac{t x^{\star T}\bar{\lambda}}{\sqrt{x^{\star T}\bar{B}x^\star}} = -\infty.$$

However, this is contradictory to the fact that

$$\frac{x^{\star T}(a^\star + \lambda^\star)}{\sqrt{x^{\star T}B^\star x^\star}} = \inf_{(a,B)\in\mathcal{U},\, \lambda^*\in\mathcal{X}^*} \frac{x^{\star T}(a+\lambda)}{\sqrt{x^{\star T}Bx^\star}}$$

must be finite.

We complete the proof by showing $\lambda^{\star T}x^\star = 0$. Since $0 \in \mathcal{X}^*$, the saddle-point property (43) implies that

$$\frac{x^{\star T}(a^\star + \lambda^\star)}{\sqrt{x^{\star T}B^\star x^\star}} \leq \frac{x^{\star T}a^\star}{\sqrt{x^{\star T}B^\star x^\star}},$$

which means $x^{\star T}\lambda^\star \leq 0$. Since $\lambda \in \mathcal{X}^*$ and $x^\star \in \mathcal{X}$, we also have $x^{\star T}\lambda^\star \geq 0$.

A.3. Proof of Proposition 2. Let γ be the optimal value of (3):

$$\gamma = \sup_{x\in\mathcal{X}} \inf_{(a,B)\in\mathcal{U}} \frac{x^T a}{\sqrt{x^T Bx}}. \tag{51}$$

We can see that for any $x \in \mathcal{X}$, the set $X = \{(\sqrt{x^T Bx}, x^T a) \mid (a, B) \in \mathcal{U}\}$ cannot lie entirely above the line $r = \gamma\sigma$ in the (σ, r) space.

Using the Cauchy–Schwarz inequality, we can show that for any nonzero x and y

$$\frac{1}{2}\left(\sqrt{x^T Bx} + \sqrt{y^T By}\right) \geq \left(\left(\frac{x+y}{2}\right)^T B \left(\frac{x+y}{2}\right)\right)^{1/2}. \tag{52}$$

Here equality holds if and only if x and y are linearly dependent.

Suppose that there are two solutions $x^\star$ and $y^\star$ which are not linearly dependent. Then, the two sets

$$X = \left\{\left(\sqrt{x^{\star T}Bx^\star}, x^{\star T}a\right) \,\middle|\, (a, B) \in \mathcal{U}\right\}, \quad Y = \left\{\left(\sqrt{y^{\star T}By^\star}, y^{\star T}a\right) \,\middle|\, (a, B) \in \mathcal{U}\right\}$$

lie on and above, but cannot lie entirely above, the line $r = \gamma\sigma$ in the (σ, r) space. If $x^\star$ and $y^\star$ are not linearly dependent, then it follows from (52) and the compactness of $\mathcal{U}$ that the set $Z = \{(\sqrt{z^{\star T}Bz^\star}, z^{\star T}a) \mid (a, B) \in \mathcal{U}\}$, with $z^\star = (x^\star + y^\star)/2$, lies entirely above the line $r = \gamma\sigma$. Therefore, we have

$$\inf_{(a,B)\in\mathcal{U}} \frac{z^{\star T}a}{\sqrt{z^{\star T}Bz^\star}} > \gamma,$$

which is contradictory to the definition of γ given in (51).

Acknowledgments. The authors thank Alessandro Magnani, Almir Mutapcic, and Young-Han Kim for helpful comments and suggestions.

REFERENCES

[1] S. Benson and Y. Ye, *DSDP5: Software for Semidefinite Programming*, http://www-unix.mcs.anl.gov/DSDP/(2005).

[2] D. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Cambridge, MA, 2003.

[3] D. Bertsimas and I. Popescu, *Optimal inequalities in probability theory: A convex optimization approach*, SIAM J. Optim., 15 (2005), pp. 780–804.

[4] M. Best and P. Grauer, *On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results*, Rev. Financ. Stud., 4 (1991), pp. 315–342.

[5] F. Black and R. Litterman, *Global portfolio optimization*, Financ. Analysts J., 48 (1992), pp. 28–43.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, 1994.

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

[8] J. Cochrane, *Asset Pricing*, 2nd ed., Princeton University Press, Princeton, NJ, 2001.

[9] O. Costa and J. Paiva, *Robust portfolio selection using linear matrix inequalities*, J. Econom. Dynam. Control, 26 (2002), pp. 889–909.

[10] L. E. Ghaoui, M. Oks, and F. Oustry, *Worst-case value-at-risk and robust portfolio optimization: A conic programming approach*, Oper. Res., 51 (2003), pp. 543–556.

[11] D. Goldfarb and G. Iyengar, *Robust portfolio selection problems*, Math. Oper. Res., 28 (2003), pp. 1–38.

[12] B. Halldórsson and R. Tütüncü, *An interior-point method for a class of saddle point problems*, J. Optim. Theory Appl., 116 (2003), pp. 559–590.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Ser. Statist., Springer-Verlag, New York, 2001.

[14] P. Jorion, *Bayes-Stein estimation for portfolio analysis*, J. Financ. Quant. Anal., 21 (1979), pp. 279–292.

[15] S. Kassam and V. Poor, *Robust techniques for signal processing: A survey*, Proc. IEEE, 73 (1985), pp. 433–481.

[16] S.-J. Kim and S. Boyd, *Two-fund separation under model mis-specification*, manuscript. Available from http://www.stanford.edu/~boyd/rob_two_fund_sep.html.

[17] S.-J. Kim, A. Magnani, and S. Boyd, *Robust Fisher discriminant analysis*, in Adv. Neural Inform. Process. Syst., MIT Press, Cambridge, MA, 2006.

[18] M. Lobo and S. Boyd, *The worst-case risk of a portfolio*, manuscript. Available from http://faculty.fuqua.duke.edu/2000.

[19] D. Luenberger, *Investment Science*, Oxford University Press, New York, 1998.

[20] H. Markowitz, *Portfolio selection*, J. Finance, 7 (1952), pp. 77–91.

[21] R. Michaud, *The Markowitz optimization enigma: Is 'optimized' optimal?*, Financ. Analysts J., 45 (1989), pp. 31–42.

[22] A. Mutapcic, S.-J. Kim, and S. Boyd, *Array signal processing with robust rejection constraints via second-order cone programming*, in Proceedings of the 40th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, 2006.

[23] A. Roy, *Safety first and the holding of assets*, Econometrica, 20 (1952), pp. 413–449.

[24] W. Sharpe, *The Sharpe ratio*, J. Portfolio Manag., 21 (1994), pp. 49–58.

[25] M. Sion, *On general minimax theorems*, Pacific J. Math., 8 (1958), pp. 171–176.

[26] J. Sturm, *Using SEDUMI 1.02, a Matlab Toolbox for Optimization Over Symmetric Cones*, http://www.fewcal.kub.nl/sturm/software/sedumi.html/(2001).

[27] K. Toh, R. H. Tütüncü, and M. Todd, *SDPT3 version 3.02. A Matlab software for semidefinite-quadratic-linear programming*, http://www.math.nus.edu.sg/~mattohkc/sdpt3.html/ (2002).

[28] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley-Interscience, New York, 2001.

[29] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.

[30] L. Vandenberghe, S. Boyd, and K. Comanor, *Generalized Chebyshev bounds via semidefinite programming*, SIAM Rev., 49 (2007), pp. 52–64.

[31] S. Verdú and H. Poor, *On minimax robustness: A general approach and applications*, IEEE Trans. Inform. Theory, 30 (1984), pp. 328–340.