Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

A Minimax Theorem with Applications to Machine Learning, Signal Processing, and Finance

Seung-Jean Kim Stephen Boyd
Information Systems Laboratory
Electrical Engineering Department, Stanford University
Stanford, CA 94305-9510
sjkim@stanford.edu boyd@stanford.edu

***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, 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.**

I. Introduction

This paper concerns a fractional function of the form

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

where $x, a \in \mathbf{R}^n$ and $B = B^T \in \mathbf{R}^{n\times n}$. We assume that $x \in \mathcal{X} \subseteq \mathbf{R}^n \backslash \{0\}$ and $(a, B) \in \mathcal{U} \subseteq \mathbf{R}^n \times \mathbf{S}^n_{++}$. Here $\mathbf{S}^n_{++}$ 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 0) in x: for all $t > 0$,

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

If

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

then for fixed $(a, B) \in \mathcal{U}$, f is quasiconcave in x, and for fixed $x \in \mathcal{X}$, f is quasiconvex 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 $\mathbf{R}^n$), 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).

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:*

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

  with variables $x \in \mathbf{R}^n$.
- *Min-max problem:*

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

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

The minimax inequality or *weak minimax property*

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

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

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

holds if $\mathcal{X}$ is convex, $\mathcal{U}$ is convex and compact, and (2) holds, which follows from Sion's quasiconvex-quasiconcave minimax theorem [23].

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

Theorem 1: Suppose that $\mathcal{X}$ is a cone in $\mathbf{R}^n$, that does not contain the origin, with $\mathcal{X} \cup \{0\}$ convex and closed, and $\mathcal{U}$ is a convex compact subset of $\mathbf{R}^n \times \mathbf{S}^n_{++}$. Suppose further that

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

Consider the optimization problem

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

with variables $a \in \mathbf{R}^n$, $B = B^T \in \mathbf{R}^{n\times n}$, and $\lambda \in \mathbf{R}^n$, where $\mathcal{X}^*$ is the dual cone of $\mathcal{X}$, *i.e.*,

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

Then, this problem has a solution, say $(a^\star, B^\star, \lambda^\star)$, that satisfies

$$a^\star + \lambda^\star \neq 0, \qquad x^\star = B^{\star -1}(a^\star + \lambda^\star) \in \mathcal{X},$$

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

$$\frac{x^T a^\star}{\sqrt{x^T B^\star x}} \leq \frac{x^{\star T} a^\star}{\sqrt{x^{\star T} B^\star x^\star}} \leq \frac{x^{\star T} a}{\sqrt{x^{\star T} B x^\star}}, \qquad (9)$$
$$\forall x \in \mathcal{X}, \quad \forall (a, B) \in \mathcal{U}.$$

The proof is given in the full version [13] of the paper.

From a standard result [3, §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), and the pair $(\mu^\star, \Sigma^\star)$ solves (4).

This minimax theorem has applications in machine learning (robust Fisher linear discriminant analysis), signal processing (robust beamforming, robust matched filtering), and finance (robust portfolio selection). In these applications, x corresponds to some design or model parameters to be chosen, and the pair (a, B) corresponds to the statistical model, which is uncertain.

The paper is organized as follows. We give a probabilistic interpretation of the minimax result in Section II. We then give the applications of the minimax result in machine learning, signal processing, and finance in Section III–Section V. We give our conclusions in Section VI.

II. A PROBABILISTIC INTERPRETATION

In this section, we give a probabilistic interpretation of Theorem 1.

A. *Probabilistic linear separation*

Suppose $z \sim \mathcal{N}(a, B)$, and $x \in \mathbf{R}^n$. 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

$$\mathbf{Prob}(x^Tz \geq 0) = \Phi\left(\frac{x^Ta}{\sqrt{x^TBx}}\right), \qquad (10)$$

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

Theorem 1 with $\mathcal{U} = \{(a, B)\}$ tells us that the righthand 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 $\Phi\left(\left[(a + \lambda^\star)^TB^{-1}(a + \lambda^\star)\right]^{1/2}\right)$. Thus, $(a+\lambda^\star)^TB^{-1}(a+\lambda^\star)$ (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. Here $\mathbf{E}$ denotes the expectation operation. Then, $\mathbf{E}\,x^Tz = x^Ta$ and $\mathbf{E}(x^Tz - x^Ta)^2 = x^TBx$, so by the Chebyshev bound, we have

$$\mathbf{Prob}(x^Tz \geq 0) \geq \Psi\left(\frac{x^Ta}{\sqrt{x^TBx}}\right), \qquad (11)$$

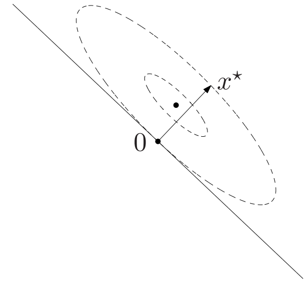

Fig. 1. Illustration of $x^\star = B^{-1}a$. The center of the two confidence ellipsoids (whose boundaries are shown as dashed curves) is a, and their shapes are determined by B.

where $\Psi(u) = \max\{u, 0\}^2/1 + \max\{u, 0\}^2$. 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) [4], [25]. 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 $\mathbf{Prob}(x^Tz \geq 0)$. The maximum value of Chebyshev lower bound is $p^\star/(1 + p^\star)$, where $p^\star = \left[(a + \lambda^\star)^TB^{-1}(a + \lambda^\star)\right]^{1/2}$. 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 $\mathcal{X} = \mathbf{R}^n\backslash\{0\}$, we have $\mathcal{X}^* = 0$, so $x = B^{-1}a$ maximizes the righthand 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 $\mathbf{R}^2$.

B. *Worst-case statistics*

We now assume that the mean and covariance are uncertain, but known to belong to a convex compact subset $\mathcal{U}$ of $\mathbf{R}^n \times \mathbf{S}^n_{++}$. We make one technical 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.

For fixed x, the problem of finding the worst-case statistics can be written as

$$\begin{array}{ll} \text{minimize} & \mathbf{Prob}(x^Tz > 0) \\ \text{subject to} & (a, B) \in \mathcal{U} \end{array}$$

where $z \sim \mathcal{N}(a, B)$ and the variables are $a \in \mathbf{R}^n$ and $B = B^T \in \mathbf{R}^{n\times n}$. This problem is equivalent to

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

C. An interpretation of the minimax result

Consider the problem of finding a hyperplane through the origin and perpendicular to $x \in \mathcal{X}$ that separates a normal random variable z on $\mathbf{R}^n$ with uncertain first and second moments belonging to $\mathcal{U}$:

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

This problem is equivalent to

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

Theorem 1 gives an effective solution method for this problem, based on convex optimization.

We close by pointing out that at the saddle point $(a^\star, B^\star, \lambda^\star)$ in Theorem 1, the pair $(a^\star, B^\star)$ solves

$$\begin{array}{ll} \text{minimize} & \sup_{x\in\mathcal{X}} x^T a/\sqrt{x^T B x} \\ \text{subject to} & (a,B) \in \mathcal{U}, \end{array}$$

so it solves

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

In other words, the pair $(a^\star, B^\star)$ gives the least favorable statistics, with x chosen optimally to maximize the separation probability.

III. Robust Fisher discriminant analysis

As another application, we consider a robust classification problem.

A. Fisher linear discriminant analysis

In linear discriminant analysis (LDA), we want to separate two classes which can be identified with two random variables in $\mathbf{R}^n$. Fisher linear discriminant analysis (FLDA) is a widely-used technique for pattern classification, proposed by R. Fisher in the 1930s. The reader is referred to standard textbooks on statistical learning, *e.g.*, [9], for more on FLDA.

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

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

where μ_+ and Σ_+ (μ_+ and Σ_-) denote the mean and covariance 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

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

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

$$\phi(x) = \mathrm{sgn}(\bar{w}^T x + v), \tag{12}$$

where

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

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

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

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

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

B. 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 [14] 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 $\mathcal{U}$ of $\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{S}^n_{++} \times \mathbf{S}^n_{++}$. We make the following assumption:

$$\text{for each } (\mu_+, \mu_-, \Sigma_+, \Sigma_-) \in \mathcal{U}, \text{ we have } \mu_+ \neq \mu_-. \tag{13}$$

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

$$\begin{array}{ll} \text{minimize} & f(w, \mu_+ - \mu_-, \Sigma_+ + \Sigma_-) \\ \text{subject to} & (\mu_+, \mu_-, \Sigma_+, \Sigma_-) \in \mathcal{U}, \end{array} \tag{14}$$

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

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

(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:

$$\begin{array}{ll} \text{maximize} & \inf_{(\mu_+,\mu_-,\Sigma_+,\Sigma_-)\in\mathcal{U}} f(w, \mu_+ - \mu_-, \Sigma_+ + \Sigma_-) \\ \text{subject to} & w \neq 0, \end{array} \quad (15)$$

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 (15) is called a *robust optimal Fisher discriminant*.

The robust robust FLDA problem (15) has the form (3) We can see from (13) that the assumption (7) holds. The robust FLDA problem can therefore be solved by using the minimax result described above.

C. Numerical example

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

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

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

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

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

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

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

$$\begin{array}{rcl} \|\Sigma_+{}^{-1/2}(\mu_+ - \bar{\mu}_+)\| & = & \|\mu_+ - \bar{\mu}_+\| = 1, \\ \|\Sigma_-{}^{-1/2}(\mu_- - \bar{\mu}_-)\| & = & \|\mu_- - \bar{\mu}_-\| = 1. \end{array}$$

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

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

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 I summarizes the results. Here, $\mathbf{P}_{\rm nom}$ is the nominal discrimination probability and $\mathbf{P}_{\rm wc}$ is the worst-case

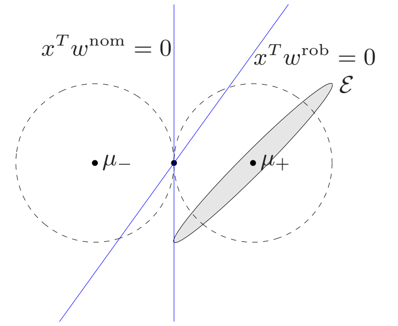

Fig. 2. A simple example for robust FLDA.

	$\mathbf{P}_{\rm nom}$	$\mathbf{P}_{\rm wc}$
nominal optimal discriminant	0.92	0.78
robust optimal discriminant	0.87	0.83

TABLE I

ROBUST DISCRIMINANT ANALYSIS RESULTS.

discrimination probability. The nominal optimal discriminant achieves $\mathbf{P}_{\rm nom} = 0.92$, 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 worst-case discrimination probability around 83%, 5% higher than that of the nominal optimal discriminant.

IV. ROBUST MATCHED FILTERING

As another application, we consider robust matched filtering, which has been extensively studied in 1980s; see, *e.g.*, [11], [10], [21], [26], [27], [28] and the survey paper [12] for robust signal processing techniques. In [28], Verdú and Poor consider a game-theoretic approach to the design of filters that are robust with respect to modeling uncertainties in the signal and covariance and describe a set of convexity and regularity conditions for the existence of a saddle point in the game when the uncertainties in the signal and covariance are separable. Most work on robust matched filtering focused on finding signal and noise covariance models which allow one to solve the robust matched filtering problem analytically not numerically. More recently, ideas from the (worst-case) robust optimization [1], [2] have been applied to robust beamforming, a special type of robust matched filtering problem [15].

In this section, we consider robust matched filtering with a general uncertainty model.

A. Matched filtering

Consider a signal model

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

where $s(t) \in \mathbf{R}$ is the desired signal, $y(t) \in \mathbf{R}^n$ is the received signal, and $v(t) \sim \mathcal{N}(0, \Sigma)$ is the noise. The filtered output with weight vector $w \in \mathbf{R}^n$ is given by

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

The goal is to detect the presence of the desired signal (which usually takes its value from $\{0, 1\}$).

In (standard) matched filtering, we want to choose w that maximizes the signal to noise ratio (SNR):

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

This problem is equivalent to maximizing the square root of the SNR (SSNR)

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

The filter, called the *nominal optimal filter*, that maximizes SSNR, is given by $w = \Sigma^{-1} a$. When the covariance is an identity matrix, the matched filter $w = a$ is optimal. (See, *e.g.*, [24] for more on matched filtering.)

Once the filter coefficients are found, we can use a simple thresholding rule

$$h(a) = \begin{cases} \text{there is no signal}, & w^T a < t \\ \text{there is a signal}, & w^T a > t, \end{cases} \tag{16}$$

to detect whether the desired signal is present. Here, t is the threshold. By varying the threshold over $\mathbf{R}$, we can obtain the optimal receiver operating characteristic (ROC) curve, which describes a fundamental limit of detection performance [24].

B. Robust matched filtering

We assume that the steering vector and covariance matrix are uncertain, but known to belong to a convex compact subset $\mathcal{U}$ of $\mathbf{C}^n \times \mathbf{S}^n_{++}$. We make one technical assumption: for each pair $(a, \Sigma) \in \mathcal{U}$, we have $a \neq 0$.

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

$$\begin{array}{ll} \text{minimize} & f(w, a, \Sigma) \\ \text{subject to} & (a, \Sigma) \in \mathcal{U}, \end{array} \tag{17}$$

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. This problem can be cast as the optimization problem

$$\begin{array}{ll} \text{maximize} & \inf_{(a,\Sigma)\in\mathcal{U}} f(x, a, B) \\ \text{subject to} & w \neq 0, \end{array} \tag{18}$$

with variables w. (The thresholding rule (16) that uses a solution of this problem as the weight vector yields the robust ROC curve that specifies limits of performance in the worst-case sense.) This problem can be solved using Theorem 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 [19]. This problem can also be solved using Theorem 1.

	nominal SSNR	worst-case SSNR
nominal optimal filter	5.5	3.0
robust optimal filter	4.9	3.6

TABLE II

ROBUST MATCHED FILTERING RESULTS.

C. 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

$$\left[\begin{array}{cccc} 1 & - & + & - \\ & 1 & ? & + \\ & & 1 & ? \\ & & & 1 \end{array}\right].$$

(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

$$\bar{\Sigma} = \left[\begin{array}{cccc} 1 & -.5 & .5 & -.5 \\ & 1 & 0 & .5 \\ & & 1 & 0 \\ & & & 1 \end{array}\right].$$

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

$$\Sigma^{\text{lf}} = \left[\begin{array}{cccc} 1 & 0 & .38 & -.12 \\ & 1 & .41 & .74 \\ & & 1 & .23 \\ & & & 1 \end{array}\right].$$

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

Table II summarizes the results. The nominal optimal filter achieves SSNR 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 3.6, which is 20% larger than that of the nominal optimal filter.

V. WORST-CASE SHARPE RATIO MAXIMIZATION

As the final application, we consider robust asset allocation.

A. Mean-variance asset allocation

Since the pioneering work of Markowitz [17], [18], 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 as a random vector $a = (a_1, \ldots, a_n)$ in $\mathbf{R}^n$. The data or asset statistics needed for MV analysis are the mean μ and covariance matrix Σ of a:

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

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 random variable $w^T a = \sum_{i=1}^n w_i a_i$ whose mean and volatility (standard deviation) are $\mu^T w$ and $\sqrt{w^T \Sigma w}$.

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. We assume that an admissible portfolio $w = (w_1, \ldots, w_n)$ is constrained to lie in a convex compact subset $\mathcal{A}$ of $\mathbf{R}^n$. The set of admissible portfolios under the portfolio budget constraint is given by

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

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

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

so the probability of outperforming a risk-free asset with return $\mu_{\rm rf}$ is

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

This probability is maximized by $w \in \mathcal{W}$ that maximizes the *Sharpe ratio (SR)*

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

This portfolio is called the *tangency portfolio (TP)*. (See, *e.g.*, [5], [22] for more on the role of the tangency portfolio in asset pricing theory and practice.)

Suppose $\mathbf{E}\, a = \mu$, $\mathbf{E}\,(a-\mu)^T(a-\mu) = \Sigma$ and otherwise arbitrary. Then, $\mathbf{E}\, a^T w = \mu^T x$, $\mathbf{E}\,(a^T x - \mu^T x)^2 = w^T\Sigma w$, so it follows from the Chebyshev bound that

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

This bound is also maximized by the tangency portfolio, since Ψ is increasing.

B. Robust allocation

In 'standard' MV portfolio analysis, we assume that the input parameters, namely, the mean vector and covariance matrix of asset returns, are known for certain. In practice, the input parameters are estimated through a statistical estimation procedure and hence uncertain due to the imperfections in the estimation procedure. Standard MV analysis is often sensitive to uncertainty or estimation error in the parameters, meaning that MV efficient 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.

There has been a growing interest in (worst-case) robust MV analysis and optimization as a systematic way of finding portfolio weights that not only work well with the particular baseline or nominal model, but that also performs reasonably well despite model uncertainty or mis-specification; see, *e.g.*, [6], [7], [8], [16], [20]. The basic idea is to explicitly incorporate a model of data uncertainty in the formulation of a portfolio optimization problem, and to optimize for the worst-case scenario under this model.

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 $\mathbf{R}^n \times \mathbf{S}^n_{++}$:

$$(\mu,\Sigma) \in \mathcal{U} \subset \mathbf{R}^n \times \mathbf{S}^n_{++}.$$

Here, we assume that

$$\begin{array}{l} \text{there exists a portfolio } \bar{w} \in \mathcal{W} \\ \text{such that } \mu^T w > \mu_{\rm rf} \text{ for all } (\mu,\Sigma) \in \mathcal{U}. \end{array} \tag{19}$$

For fixed w, the *worst-case SR analysis problem* can be formulated as

$$\begin{array}{ll} \text{minimize} & S(w,\mu,\Sigma) \\ \text{subject to} & (\mu,\Sigma) \in \mathcal{U}, \end{array} \tag{20}$$

in which the optimization variables μ and Σ. Note that w is fixed here. The optimal value of this problem is called the *worst-case Sharpe ratio* (over the uncertainty set $\mathcal{U}$), and denoted as $S_{\rm wc}(w)$. In the sequel, we are only interested in case when the maximum Sharpe ratio is greater than zero.

The problem of finding a portfolio w consisting only of risky assets that maximizes the worst-case SR with the given model $\mathcal{U}$ of uncertainty can be formulated as

$$\begin{array}{ll} \text{maximize} & \inf_{(\mu,\Sigma)\in\mathcal{U}} S(w,\mu,\Sigma) \\ \text{subject to} & w \in \mathcal{W}, \end{array} \tag{21}$$

with variable $w \in \mathbf{R}^n$. This problem is called the *worst-case SR maximization problem*. The optimal solution of the problem is called a *(worst-case) robust tangency portfolio* $w_{\rm rtp}$.

We cannot apply Theorem 1 immediately to the worst-case SR maximization problem (21), since the objective is not a fractional function of the form (1). However, we can easily reformulate the problem to have our form. Note that $\mathbf{1}^T w = 1$ for all $w \in \mathcal{W}$. Therefore, the SR has the form (1) when the domain is restricted to $\mathcal{W}$:

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

The set

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

where $\mathbf{cl}\, A$ means the closure of the set A and $A\backslash B$ means the complement of B in A, is a cone in $\mathbf{R}^n$, with $\mathcal{X} \cup$

$\{0\}$ closed and convex. The assumption (19) 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 a problem of the form

$$\begin{array}{ll} \text{maximize} & \inf_{(a,\Sigma)\in\mathcal{U}} f(x, \mu - \mu_{\rm rf}\mathbf{1}, \Sigma) \\ \text{subject to} & x \in \mathcal{X}. \end{array} \quad (22)$$

This problem has a solution, say $x^\star$. It satisfies $\mathbf{1}^T x^\star \geq 0$.

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

$$w^\star = (1/\mathbf{1}^T x^\star)x^\star$$

satisfies the budget constraint and is admissible (*i.e.*, $w^\star \in \mathcal{W}$). Therefore, it is a solution to the worst-case SR maximization (21). 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 (21) has no solution, so the game involving the SR has no saddle point. The details are given in the full version [13] of this paper.

C. Numerical example

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

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

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

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

$$\bar{\Omega} = \begin{bmatrix} 1.00 & .07 & -.12 & .43 & -.11 & .44 & .25 \\ & 1.00 & .73 & -.14 & .39 & .28 & .10 \\ & & 1.00 & .14 & .5 & .52 & -.13 \\ & & & 1.00 & .04 & .35 & .38 \\ & & & & 1.00 & .7 & .04 \\ & & & & & 1.00 & -.09 \\ & & & & & & 1.00 \end{bmatrix}.$$

The nominal covariance is

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

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

The mean uncertainty model used in our study is

$$\begin{array}{rcl} |\mu_i - \bar{\mu}_i| & \leq & 0.3|\bar{\mu}_i|, \quad i = 1, \ldots, 7, \\ |\mathbf{1}^T\mu - \mathbf{1}^T\bar{\mu}| & \leq & 0.15|\mathbf{1}^T\bar{\mu}|, \end{array}$$

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 budget is allocated to each asset of the n assets) is at most 15%. The covariance uncertainty model used in our study is

$$\begin{array}{rcl} |\Sigma_{ij} - \bar{\Sigma}_{ij}| & \leq & 0.3|\bar{\Sigma}_{ij}|, \quad i, j = 1, \ldots, 7, \\ \|\Sigma - \bar{\Sigma}\|_F & \leq & 0.15\|\bar{\Sigma}\|_F. \end{array}$$

(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:

$$\begin{array}{rcl} w & = & w_{\rm long} - w_{\rm short}, \\ w_{\rm long}, w_{\rm short} & \succeq & 0, \\ \mathbf{1}^T w_{\rm short} & \leq & \eta \mathbf{1}^T w_{\rm long}, \end{array}$$

where η is a positive constant, and $w_{\rm long}$ and $w_{\rm short}$ represent the total long position and short position 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

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

where

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

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

$$\mathcal{X}^* = \left\{ \lambda \in \mathbf{R}^n \,\middle|\, A^T y + \begin{bmatrix} \lambda \\ -\lambda \end{bmatrix} = 0 \text{ for some } y \succeq 0 \right\}.$$

We can find the robust tangency portfolio, by applying Theorem 1 to the corresponding problem (22) with the asset allocation constraints and uncertainty model described above. The nominal tangency portfolio can be found using Theorem 1 with the singleton $\mathcal{U} = \{(\bar{\mu}, \bar{\Sigma})\}$.

Table III shows the nominal and worst-case SR 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 robust market portfolio decreases about 70% from 0.74 to 0.22.

Table IV 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

	nominal SR	worst-case SR
nominal TP	0.74	0.22
robust TP	0.57	0.36

TABLE III

NOMINAL AND WORST-CASE SR OF NOMINAL AND ROBUST TANGENCY PORTFOLIOS.

	$\mathbf{P}_{\text{nom}}$	$\mathbf{P}_{\text{wc}}$
nominal TP	0.77	0.59
robust TP	0.71	0.64

TABLE IV

OUTPERFORMANCE PROBABILITY OF NOMINAL AND ROBUST TANGENCY PORTFOLIOS.

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.

VI. 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 $A \in \mathbf{R}^{n\times n}$ and $B \in \mathbf{R}^{n\times n}$ evaluated at $x \in \mathbf{R}^n$, is not possible; see, *e.g.*, [28] for a counterexample. However, the arguments can be extended to the special case when A is a dyad, *i.e.*, $A = aa^T$ with $a \in \mathbf{R}^n$, and $\mathcal{X} = \mathbf{R}^n\backslash\{0\}$. In this case, the minimax equality

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

holds with the assumption (7); see [14] for the proof.

Acknowledgments

This material is based upon work supported by the Focus Center Research Program Center for Circuit & System Solutions award 2003-CT-888, by JPL award I291856, by the Precourt Institute on Energy Efficiency, by Army award W911NF-07-1-0029, by NSF award ECS-0423905, by NSF award 0529426, by DARPA award N66001-06-C-2021, by NASA award NNX07AEIIA, by AFOSR award FA9550-06-1-0514, and by AFOSR award FA9550-06-1-0312.

REFERENCES

[1] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
[2] A. Ben-Tal and A. Nemirovski. Robust optimization: methodology and applications. *Mathematical Programming Series B*, 92(3):453–480, May 2002.
[3] D. Bertsekas, A. Nedić, and A. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.
[4] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.
[5] J. Cochrane. *Asset Pricing*. Princeton University Press, NJ: Princeton, second edition, 2001.
[6] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case Value-At-Risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.
[7] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.
[8] B. Halldórsson and R. Tütüncü. An interior-point method for a class of saddle point problems. *Journal of Optimization Theory and Applications*, 116(3):559–590, 2003.
[9] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer-Verlag, New York, 2001.
[10] S. Kassam and T. Lim. Robust Wiener filters. *Journal of Franklin Institute*, 304:171–185, 1977.
[11] S. Kassam, T. Lim, and L. Climini. Two dimensional filters for signal processing under modeling uncertainties. *IEEE Transactions on Georscience Electronics*, 18:331–336, 1980.
[12] S. Kassam and V. Poor. Robust techniques for signal processing: A survey. *Proceedings of IEEE*, 73:433–481, 1985.
[13] S.-J. Kim and S. Boyd. A minimax theorem with applications to machine learning, signal processing, and finance, 2006. Manuscript. Available from www.stanford.edu/~boyd/minimax_frac.html.
[14] S.-J. Kim, A. Magnani, and S. Boyd. Robust Fisher discriminant analysis. In *Advances in Neural Information Processing Systems*, MIT Press, 2006.
[15] S.-J. Kim, A. Magnani, A. Mutapcic, S. Boyd, and Z.-Q. Luo. Robust beamforming via worst-case SINR maximization, 2007. Accepted to *IEEE Transactions on Signal Processing*. Available from www.stanford.edu/~boyd/robust_bf.html.
[16] M. Lobo and S. Boyd. The worst-case risk of a portfolio. Unpublished manuscript. Available from http://faculty.fuqua.duke.edu/%7Emlobo/bio/researchfiles/rsk-bnd.pdf, 2000.
[17] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.
[18] H. Markowitz. *Portfolio Selection*. J. Wiley & Sons, New York, 1959.
[19] 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, October 2006.
[20] M. Pinar and R. Tütüncü. Robust profit opportunities in risky financial portfolios. *Operations Research Letters*, 33(4):331–340, 2005.
[21] H. Poor. Robust matched filters. *IEEE Transactions on Information Theory*, 29(5):677–687, 1983.
[22] W. Sharpe. The Sharpe ratio. *Journal of Portfolio Management*, 21(1):49–58, 1994.
[23] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8:171–176, 1958.
[24] H. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley-Interscience, 2001.
[25] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.
[26] S. Verdú and H. Poor. Minimax robust discrete-time matched filters. *IEEE Transactions on Communications*, 31(2):208–215, 1983.
[27] S. Verdú and H. Poor. Signal selection for robust matched filtering. *IEEE Transactions on Communications*, 31(5):667–670, 1983.
[28] S. Verdú and H. Poor. On minimax robustness: A general approach and applications. *IEEE Transactions on Information Theory*, 30(2):328–340, 1984.