# QUADRATIC AND CONVEX MINIMAX CLASSIFICATION PROBLEMS

Tomonari Kitahara        Shinji Mizuno        Kazuhide Nakata
*Tokyo Institute of Technology*

*Abstract*      When there are two classes whose mean vectors and covariance matrices are known, Lanckriet et al. [7] consider the Linear Minimax Classification (LMC) problem and they propose a method for solving it. In this paper we first discuss the Quadratic Minimax Classification (QMC) problem, which is a generalization of LMC. We show that QMC is transformed to a parametric Semidefinite Programming (SDP) problem. We further define the Convex Minimax Classification (CMC) problem. Though the two problems are generalizations of LMC, we prove that solutions of these problems can be obtained by solving LMC.

**Keywords**: Optimization, minimax classification, semidefinite programming, convexity

## 1.   Introduction

Classification is a widely used technique in real life. Its practical importance has attracted a large number of researches. These researches include, for example, classification using the Mahalanobis distance, the Neural Network, and the Support Vector Machine (SVM).

Recently, Lanckriet et al. [7] propose a minimax approach to binary classification. Their approach needs mean vectors and covariance matrices of two classes, and no assumptions are made with respect to the probability distributions. Generally, the performance of a classifier is affected by the unknown distributions. To avoid the poor performance of a classifier, Lanckriet et al. [7] propose to use a linear classifier which minimizes the worst-case misclassification probability under all the possible choice of probability distributions with given mean vectors and covariance matrices. In this paper, we call this problem the Linear Minimax Classification (LMC) problem. Lanckriet et al. [7] show that LMC is formulated as a Second Order Cone Programming (SOCP) problem. The size of this problem depends only on the dimension and it is independent from the number of data, so the problem is easily solved by an interior point method. Their algorithm is called the Minimax Probability Machine (MPM). They demonstrate that a classifier obtained by MPM works well to practical problems. From its theoretical and practical importance, many researches related to MPM have been made. Huang et al. [4] propose the Biased-MPM, which takes into account an importance of each class. In [5], the same authors as [4] further generalize the Biased-MPM and develop the Minimum Error MPM, which brings tighter worst-case accuracy. Hoi et al. [3] and Kitahara et al. [6] both extend MPM to multiple classification. Hoi et al. [3] adopt a classical method which uses binary classifiers. In contrast, Kitahara et al. [6] formulate a problem to find linear functions whose worst-case misclassification probability is small.

In most cases, classification methods are first developed in linear classification. Then they are extended to improve the performances by allowing nonlinear classification functions. For example, SVM is first proposed to give linear classifiers, after that the Kernel method

is developed for handling nonlinear classification functions. As for MPM, there are few researches in this direction. Cooke [2] analyzes quadratic classifiers which minimize the worst-case misclassification probability. But his analysis is rather restrictive due to the assumption that the covariance matrices are diagonal.

In this paper, we formally address the problem using quadratic classifiers with only the assumption that the covariance matrices are positive definite. Practically, this assumption is acceptable. We call the problem the Quadratic Minimax Classification (QMC) problem. We show that QMC is formulated as a parametric Semidefinite Programming (SDP) problem. The parametric SDP problem can be solved by an interior point method. However, in our preliminary computational experiments we observe that solutions of various examples of QMC are always linear. Motivated by these facts, we conduct a theoretical analysis of QMC.

In our analysis, we define another generalization of LMC. In both linear and quadratic classification, we essentially classify samples using a set defined by a classifier and its complementary set. Noticing this fact, we consider the problem of finding a convex set which minimizes the worst-case misclassification probability. We refer to this problem as the Convex Minimax Classification (CMC) problem. As opposed to LMC or QMC, so far we have not found any computational method to solve CMC directly. In this paper, we prove that any optimal solution of LMC is an optimal solution of CMC. Moreover, from this fact we manage to show that any optimal solution of LMC is an optimal solution of QMC.

The rest of this paper is organized as follows. In Section 2, we review LMC by Lanckriet et al. [7]. In Section 3, we define QMC and we give an SDP representation of QMC. In Section 4, we introduce CMC. Relations between the three problems are investigated in Section 5. Finally in Section 6, we conclude the paper.

Throughout this paper, we use the following notations. Let $\Re^n$ be the $n$-dimensional Euclidean space and $S^n$ be the set of $n$-dimensional symmetric matrices. $S^n_+$ and $S^n_{++}$ represent the set of $n$-dimensional symmetric positive semidefinite and positive definite matrices, respectively. For $X \in S^n$, $X \succeq O$ means that $X$ is positive semidefinite.

## 2. The Linear Minimax Classification Problem

In this section, we review the Linear Minimax Classification (LMC) problem by Lanckriet et al. [7].

Let $x \in \Re^n$ be a sample which belongs to one of two classes, namely, Class 1 or Class 2. Suppose that we know the mean vector $\mu_i \in \Re^n$ and the covariance matrix $\Sigma_i \in S^n_{++}$ of each class $i \in \{1, 2\}$. In this paper, we assume that the covariance matrices are positive definite. This assumption is valid when a regularization term is added to each of the covariance matrix. No further assumptions are made with respect to the probability distributions. We remark that the settings stated in this paragraph are valid throughout this paper.

We determine a linear classifier $l(z) = a^T z + b$, where $a \in \Re^n \setminus \{0\}$, $b \in \Re$, and $z \in \Re^n$. For any given sample $x$, if $a^T x + b < 0$ then it is classified as Class 1, if $a^T x + b > 0$ then it is classified as Class 2, and if $a^T x + b = 0$ then it can be classified as either Class 1 or Class 2.

For the classifier $l(z) = a^T z + b$, its worst-case misclassification probability of Class 1 sample as Class 2 is expressed as

$$\sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{a^T x + b \geq 0\},$$

where the supremum is taken over all probability distributions having the mean vector

$\mu_1$ and the covariance matrix $\Sigma_1$. Considering the other case similarly, the worst-case misclassification probability $\alpha_l$ of the classifier $l(z) = a^T z + b$ is given by

$$\alpha_l = \max\{ \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{a^T x + b \geq 0\}, \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{a^T x + b \leq 0\}\}. \tag{1}$$

DEFINITION 2.1 *(LMC) We refer to the problem of finding a linear function $l(z) = a^T z + b$, which minimizes the worst-case misclassification probability defined by (1), as the Linear Minimax Classification (LMC) problem. We represent the optimal value of LMC as $\alpha_1$, that is,*

$$\alpha_1 = \min_l \alpha_l. \tag{2}$$

From (1), it is easy to see that LMC can be expressed as

$$\begin{array}{ll} \min & \alpha \\ \text{subject to} & \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{a^T x + b \geq 0\} \leq \alpha, \\ & \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{a^T x + b \leq 0\} \leq \alpha, \end{array} \tag{3}$$

where $\alpha \in [0, 1]$, $a \in \Re^n \setminus \{0\}$, and $b \in \Re$ are variables. Lanckriet et al. [7] prove that if $\mu_1 \neq \mu_2$ then $\alpha_1 < 1$ and $a \neq 0$ at any optimal solution $(\alpha, a, b)$ of (3). In the following, we assume $\mu_1 \neq \mu_2$.

Lanckriet et al. [7] show that (3) is reduced to the following problem:

$$\begin{array}{ll} \min & \|\Sigma_1^{1/2} a\| + \|\Sigma_2^{1/2} a\| \\ \text{subject to} & (\mu_2 - \mu_1)^T a = 1, \end{array} \tag{4}$$

where $\Sigma_i^{1/2}$ $(i = 1, 2)$ is a symmetric positive definite matrix which satisfies $(\Sigma_i^{1/2})^2 = \Sigma_i$. The problem (4) is a Second Order Cone Programming (SOCP) problem and can be solved efficiently by an interior point method. The algorithm is called the Minimax Probability Machine (MPM). In [7], it is demonstrated that the performance of MPM is competitive to the Support Vector Machine (SVM).

## 3. The Quadratic Minimax Classification Problem

In this section, we define the Quadratic Minimax Classification (QMC) problem and present an SDP representation of it.

Under the same settings as Section 2, we consider to classify a sample using a quadratic function of the form $q(z) = z^T A z + 2b^T z + c$ $((A, b, c) \in S^n \times \Re^n \times \Re)$ in analogy with linear classification. Then, the worst-case misclassification probability $\alpha_q$ of $q(z)$ is given by

$$\alpha_q = \max\{ \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x^T A x + 2b^T x + c \geq 0\}, \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x^T A x + 2b^T x + c \leq 0\}\}. \tag{5}$$

DEFINITION 3.1 *(QMC) We call the problem of finding a quadratic function $q(z) = z^T A z + 2b^T z + c$, which minimizes the worst-case misclassification probability defined by (5), the Quadratic Minimax Classification (QMC) problem. The optimal value of QMC is represented as $\alpha_2$, that is,*

$$\alpha_2 = \min_q \alpha_q. \tag{6}$$

Since a linear function is a special quadratic function, we have

$$\alpha_2 \leq \alpha_1. \tag{7}$$

QMC can be formulated as

$$
\begin{array}{ll}
\min & \alpha \\
\text{subject to} & \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x^T A x + 2b^T x + c \geq 0\} \leq \alpha, \\
& \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x^T A x + 2b^T x + c \leq 0\} \leq \alpha,
\end{array}
\tag{8}
$$

where $\alpha \in [0, 1]$, $A \in S^n$, $b \in \Re^n$, and $c \in \Re$ are variables.

The next lemma gives an SDP representation of QMC.

LEMMA 3.1 *Suppose that $\mu_1 \neq \mu_2$. Then the optimal value of (8) is less than 1. Furthermore, the problem (8) is equivalent to the following parametric semidefinite programming problem:*

$$
\begin{array}{ll}
\min & \alpha \\
\text{subject to} & \mathbf{tr}((\Sigma_i + \mu_i \mu_i^T) P_i) + 2\mu_i^T q_i + r_i \leq \tau_i \alpha \ (i = 1, 2), \\
& \begin{bmatrix} P_i & q_i \\ q_i^T & r_i \end{bmatrix} \succeq O \ (i = 1, 2), \\
& \begin{bmatrix} P_1 - A & q_1 - b \\ (q_1 - b)^T & r_1 - \tau_1 - c \end{bmatrix} \succeq O, \\
& \begin{bmatrix} P_2 + A & q_2 + b \\ (q_2 + b)^T & r_2 - \tau_2 + c \end{bmatrix} \succeq O, \\
& \tau_i > 0 \ (i = 1, 2),
\end{array}
\tag{9}
$$

*where $\alpha, (P_i, q_i, r_i, \tau_i) \in S^n \times \Re^n \times \Re \times \Re$ $(i = 1, 2)$, and $(A, b, c) \in S^n \times \Re^n \times \Re$ are variables.*

The following lemma is needed in the proof of Lemma 3.1.

LEMMA 3.2 *The value $\sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x^T A x + 2b^T x + c \geq 0\}$ is the optimal value of the following semidefinite programming problem:*

$$
\begin{array}{ll}
\min & \mathbf{tr}((\Sigma_1 + \mu_1 \mu_1^T) P_1') + 2\mu_1^T q_1' + r_1' \\
\text{subject to} & \begin{bmatrix} P_1' & q_1' \\ (q_1')^T & r_1' \end{bmatrix} \succeq O, \\
& \begin{bmatrix} P_1' - \tau_1' A & q_1' - \tau_1' b \\ (q_1' - \tau_1' b)^T & r_1' - 1 - \tau_1' c \end{bmatrix} \succeq O, \\
& \tau_1' \geq 0,
\end{array}
\tag{10}
$$

*where $(P_1', q_1', r_1', \tau_1') \in S^n \times \Re^n \times \Re \times \Re$ are variables.*

PROOF.   The proof is given in Vandenberghe et al. [10]. Here we give a sketch of the proof.

Define the set

$$D = \{x \in \Re^n : x^T A x + 2b^T x + c \geq 0\}.$$

From the constraints of (10), for any $x \in \Re^n$ we have

$$
\begin{array}{l}
x^T P_1' x + 2q_1'^T x + r_1' \geq 1 + \tau_1'(x^T A x + 2b^T x + c), \\
x^T P_1' x + 2q_1'^T x + r_1' \geq 0.
\end{array}
$$

Thus it follows that

$$x^T P_1' x + 2q_1'^T x + r_1' \geq \delta_D(x) = \begin{cases} 1; & x \in D \\ 0; & x \notin D \end{cases}$$

Therefore, for $x \sim (\mu_1, \Sigma_1)$, we have

$$\begin{aligned} \mathbf{tr}((\Sigma_1 + \mu_1 \mu_1^T) P_1') + 2\mu_1^T q_1' + r_1' &= \mathrm{E}[x^T P_1' x + 2q_1'^T x + r_1'] \\ &\geq \mathrm{E}[\delta_D(x)] = \Pr\{x \in D\}. \end{aligned}$$

The above relation implies that the optimal value of (10) is an upper bound for the value $\sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x^T A x + 2b^T x + c \geq 0\}$. We can also show that the optimal value of the dual problem for (10) gives a lower bound. Notice that (10) is strictly feasible. From these facts and the semidefinite programming duality, the statement follows. □

PROOF OF LEMMA 3.1. As we mentioned in Section 1, when $\mu_1 \neq \mu_2$, we have $\alpha_1 < 1$, which implies $\alpha_2 < 1$.

To show the second statement, we claim that when $\alpha < 1$, the first constraint of (8) is equivalent to the existence of $(P_1, q_1, r_1, \tau_1)$ which satisfies the constraints of (9).

From Lemma 3.2, the constraint is equivalent to the existence of $(P_1', q_1', r_1', \tau_1') \in \mathrm{S}^n \times \Re^n \times \Re \times \Re$ satisfying the following system:

$$\mathbf{tr}((\Sigma_1 + \mu_1 \mu_1^T) P_1') + 2\mu_1^T q_1' + r_1' \leq \alpha, \tag{11}$$

$$\begin{bmatrix} P_1' & q_1' \\ (q_1')^T & r_1' \end{bmatrix} \succeq O,$$

$$\begin{bmatrix} P_1' - \tau_1' A & q_1' - \tau_1' b \\ (q_1' - \tau_1' b)^T & r_1' - 1 - \tau_1' c \end{bmatrix} \succeq O,$$

$$\tau_1' \geq 0.$$

If $\tau_1' = 0$, then we have

$$\begin{bmatrix} P_1' & q_1' \\ (q_1')^T & r_1' - 1 \end{bmatrix} \succeq O,$$

which leads to

$$\mathbf{tr}\left( \begin{bmatrix} P_1' & q_1' \\ (q_1')^T & r_1' - 1 \end{bmatrix} \cdot \begin{bmatrix} \Sigma_1 + \mu_1 \mu_1^T & \mu_1 \\ \mu_1^T & 1 \end{bmatrix} \right) \geq 0$$

or equivalently

$$\mathbf{tr}((\Sigma_1 + \mu_1 \mu_1^T) P_1') + 2(q_1')^T \mu_1 + r_1' \geq 1,$$

which contradicts to (11). Thus we have $\tau_1' \neq 0$. Then $(P_1, q_1, r_1, \tau_1) = \frac{1}{\tau_1'}(P_1', q_1', r_1', 1)$ satisfy the constraints of (9). Similarly, we can show that the second constraint of (8) is equivalent to the existence of $(P_2, q_2, r_2, \tau_2)$ satisfying the constraints of (9).

Thus we have established the fact that when $\alpha < 1$, the constraints of (8) is equivalent to those of (9), which completes the proof. □

If we fix $\alpha$, the constraints of (9) are semidefinite constraints, and the resulting feasibility problem is an SDP problem. Furthermore, the constraints of (9) are monotonic in $\alpha$, that is, if the constraints are infeasible for some $\alpha'$, then they are infeasible for $\alpha \leq \alpha'$. Exploiting these facts, we can compute the optimal value of (9) by using a line search on $\alpha$, for example, a bisection method.

In our preliminary computational experiments, we observe that solutions of various examples of QMC are always linear. In section 5, we prove that this property holds in general.

## 4.  The Convex Minimax Classification Problem

In this section, we generalize LMC by noticing the fact that the classification rule is characterized by a set.

We define an open half space

$$S = \{z \in \Re^n | a^T z + b < 0\}$$

for a linear classifier $l(z) = a^T z + b$. Let $\text{ext}(S)$ and $\text{bd}(S)$ denote the exterior and boundary of $S$, respectively. Then the worst-case misclassification probability (1) is equivalent to

$$\alpha = \max\{ \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x \in \text{ext}(S) \cup \text{bd}(S)\}, \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x \in S \cup \text{bd}(S)\}\} . \quad (12)$$

The LMC problem is to find an open half space $S$, which minimizes $\alpha$ defined by (12).

Now we consider a classification rule using a general set which is not restricted to a half space. Suppose we have a set $S \subset \Re^n$. For any sample $x \in \Re^n$, if $x \in S$, then it is classified as Class 1, if $x \in \text{ext}(S)$ then it is classified as Class 2, and if $x \in \text{bd}(S)$, then it can be classified as either Class 1 or Class 2. As in the former sections, the worst-case misclassification probability $\alpha_S$ of $S$ is expressed as

$$\alpha_S = \max\{ \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x \in \text{ext}(S) \cup \text{bd}(S)\}, \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x \in S \cup \text{bd}(S)\}\} . \quad (13)$$

It is possible to consider $S$ to be a general set, but due to its generality, the problem is intractable. Therefore, as the first step, in this paper we focus on the case where $S$ is an open convex set.

DEFINITION 4.1 *(CMC) We call the problem of finding an open convex set $S$, which minimizes the worst-case misclassification probability defined by (13), the Convex Minimax Classification (CMC) problem. The optimal value of CMC is denoted by $\alpha_3$, that is,*

$$\alpha_3 = \min_{S:\text{open convex}} \alpha_S. \quad (14)$$

It is easy to see that

$$\alpha_3 \leq \alpha_1.$$

As in the former sections, CMC can be expressed as

$$\begin{array}{ll} \min & \alpha \\ \text{subject to} & \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x \in \text{ext}(S) \cup \text{bd}(S)\} \leq \alpha, \\ & \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x \in S \cup \text{bd}(S)\} \leq \alpha, \\ & S \subset \Re^n : \text{ open, convex.} \end{array} \quad (15)$$

In contrast to LMC or QMC, we have not found any solvable optimization problem for CMC.

## 5.  Relations between Three Problems

In this section, we investigate relations between LMC, QMC, and CMC. We state our main results as the following two theorems.

THEOREM 5.1 *Suppose that $\mu_1 \neq \mu_2$. Then any optimal solution of LMC is an optimal solution of QMC. In short, we have $\alpha_1 = \alpha_2$.*

THEOREM 5.2 *Suppose that $\mu_1 \neq \mu_2$ and let $a_*^T z + b_*$ be any optimal solution of LMC. Then the half space*

$$H^* = \{z \in \Re^n \mid a_*^T z + b_* < 0\} \tag{16}$$

*is an optimal solution of CMC, that is, $\alpha_1 = \alpha_3$.*

Remember that both QMC and CMC are generalizations of LMC, so there is a possibility that we could obtain a better classifier than the optimal one of LMC. It is interesting to see that the above two theorems exclude this possibility. We also remark that Cooke [2] shows the same result as Theorem 5.1 under the assumption that covariance matrices are diagonal. As he admits, this is a restrictive assumption. In this section, we prove the property with only the assumption that the covariance matrices are positive definite.

First we prove Theorem 5.2. Our proof heavily relies on the next classical result.

LEMMA 5.1 *(Marshall and Olkin [8]) Let $(\mu, \Sigma) \in \Re^n \times S_{++}^n$ be given. Then for any convex set $S \subset \Re^n$, it holds that*

$$\sup_{x \sim (\mu, \Sigma)} \Pr\{x \in S\} = \frac{1}{1 + d^2},$$

*where $d^2 = \inf_{y \in S}(y - \mu)^T \Sigma^{-1}(y - \mu)$.*

The result is rediscovered by Popescu [9] from an optimization perspective. Intuitively speaking, the lemma states that the supremum of the probability is determined by a kind of distance from $\mu$ to $S$.

We remark that this type of bounds, so to speak Chebyshev bounds, have regained interests with the recent developments of interior point methods for semidefinite programming problems [10]. It is because in some cases the bounds can efficiently be computed by semidefinite programming formulations, as Lemma 3.2 shows.

PROOF OF THEOREM 5.2. We prove this theorem by showing that for any open convex set $S$, there exists a half space $H$ such that $\alpha_H \leq \alpha_S$, where $\alpha_S$ and $\alpha_H$ are defined by (13).

As $S$ is convex, its closure $\mathrm{cl}(S) = S \cup \mathrm{bd}(S)$ is also convex. Thus, from Lemma 5.1, we have

$$\sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x \in \mathrm{cl}(S)\} = \frac{1}{1 + d^2},$$

where

$$d^2 = \inf_{y \in \mathrm{cl}(S)} (y - \mu_2)^T \Sigma_2^{-1} (y - \mu_2)$$

or equivalently

$$d = \inf_{y \in \mathrm{cl}(S)} \|\Sigma_2^{-1/2} y - \Sigma_2^{-1/2} \mu_2\|. \tag{17}$$

Here $\Sigma_2^{-1/2}$ is a symmetric positive definite matrix satisfying $(\Sigma_2^{-1/2})^2 = \Sigma_2^{-1}$.

The right-hand minimization problem of (17) can be seen as a projection problem. Let

$$T = \{z \in \Re^n \mid z = \Sigma_2^{-1/2} w, \ w \in \mathrm{cl}(S)\}$$

and define the problem

$$\inf_{z \in T} \|z - \Sigma_2^{-1/2} \mu_2\|. \tag{18}$$

Both Figure 1 and Figure 2 show an essence of our proof. In the transformed space shown
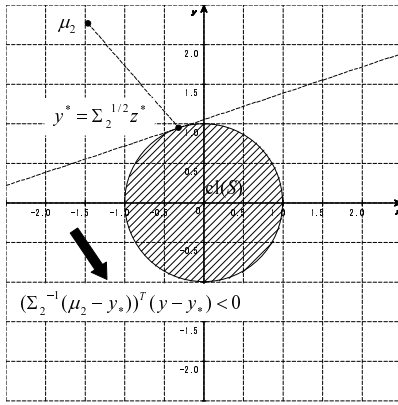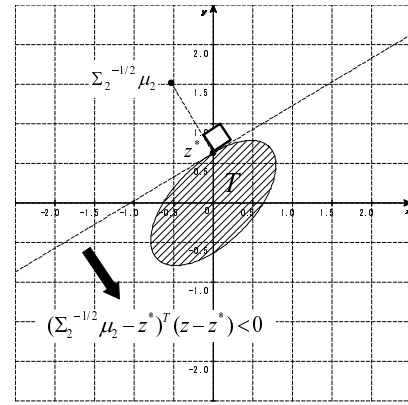
Figure 1: The original space



Figure 2: The transformed space

in Figure 2, there is a unique projection of $\Sigma^{-1/2}\mu_2$ onto $T$, which is denoted by $z^*$. Let $H'$ be a set defined by

$$H' = \{z \in \Re^n |\ (\Sigma_2^{-1/2}\mu_2 - z^*)^T(z - z^*) < 0\}.$$

Then by considering the tangent at $z^*$ it is easy to see that

$$T \subset \mathrm{cl}(H') \tag{19}$$

and

$$\inf_{z \in T} \|z - \Sigma_2^{-1/2}\mu_2\| = \inf_{z \in \mathrm{cl}(H')} \|z - \Sigma_2^{-1/2}\mu_2\| = \|z^* - \Sigma_2^{-1/2}\mu_2\|. \tag{20}$$

Back in the original space, for $y^* = \Sigma_2^{1/2}z^*$, let $H$ be a set defined by

$$H = \{y \in \Re^n |\ (\Sigma_2^{-1}(\mu_2 - y^*))^T(y - y^*) < 0\}.$$

Then (19) implies

$$S \subset H,$$

which is equivalent to

$$\mathrm{ext}(H) \cup \mathrm{bd}(H) \subset \mathrm{ext}(S) \cup \mathrm{bd}(S).$$

Thus we have

$$\sup_{x \sim (\mu_1,\Sigma_1)} \mathrm{Pr}\{x \in \mathrm{ext}(H) \cup \mathrm{bd}(H)\} \leq \sup_{x \sim (\mu_1,\Sigma_1)} \mathrm{Pr}\{x \in \mathrm{ext}(S) \cup \mathrm{bd}(S)\}. \tag{21}$$

Note that (20) shows

$$\inf_{y \in \mathrm{cl}(S)} \|\Sigma_2^{-1/2}y - \Sigma_2^{-1/2}\mu_2\| = \inf_{y \in \mathrm{cl}(H)} \|\Sigma_2^{-1/2}y - \Sigma_2^{-1/2}\mu_2\|,$$

which means

$$\sup_{x \sim (\mu_2,\Sigma_2)} \mathrm{Pr}\{x \in \mathrm{cl}(H)\} = \sup_{x \sim (\mu_2,\Sigma_2)} \mathrm{Pr}\{x \in \mathrm{cl}(S)\} \tag{22}$$

from Lemma 5.1. Combining (21) with (22), we conclude that $\alpha_H \leq \alpha_S$.

Let $a_*^T z + b_*$ be any optimal solution of LMC and define the half space $H^*$ according to (16). Obviously, $H^*$ is a feasible solution of CMC and now we have shown that its objective

value is the same as the optimal value of CMC, that is, $\alpha_1 = \alpha_{H^*} = \alpha_3$, which shows $H^*$ is an optimal solution of CMC. $\qquad\square$

Since we do not assume convexity in QMC, Theorem 5.1 is not directly derived from Theorem 5.2. But the next lemma essentially says that we can focus only on a convex quadratic classifier in QMC.

**LEMMA 5.2** *For any quadratic classifier $q(z) = z^T A z + 2b^T z + c$, there exists a convex quadratic classifier $f(z) = z^T A' z + 2b'^T z + c'$ $((A', b', c') \in \mathrm{S}^n_+ \times \Re^n \times \Re)$ such that $\alpha_f \le \alpha_q$.*

PROOF. When $\alpha_q = 1$, where $\alpha_q$ is defined by (5), any convex quadratic classifier satisfies the property of the lemma. Therefore in the following, we assume that $\alpha_q < 1$.

From Lemma 3.2, the worst-case misclassification probability of Class 1 can be computed as the optimal value of the following Semidefinite Programming (SDP) problem:

$$
\begin{aligned}
\min \quad & \mathbf{tr}((\Sigma_1 + \mu_1\mu_1^T)P) + 2\mu_1^T q + r \\
\text{subject to} \quad & \begin{bmatrix} P & q \\ q^T & r \end{bmatrix} \succeq O, \\
& \begin{bmatrix} P - \tau A & q - \tau b \\ (q - \tau b)^T & r - 1 - \tau c \end{bmatrix} \succeq O, \\
& \tau \ge 0.
\end{aligned}
$$

We denote this problem as $\mathrm{P}(A, b, c)$. Notice from the assumption $\alpha_q < 1$, the matrix

$$
\hat{A} = \begin{bmatrix} A & b \\ b^T & c \end{bmatrix}
$$

is neither positive semidefinite nor negative semidefinite. Indeed, if $\hat{A}$ were positive semidefinite, for any $x \in \Re^n$ we would have

$$
x^T A x + 2b^T x + c \ge 0,
$$

which implies

$$
\sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x^T A x + 2b^T x + c \ge 0\} = 1.
$$

Thus we would have $\alpha_q = 1$ from (5), which is a contradiction. Similarly, we can show a contradiction in the case where $\hat{A}$ were negative semidefinite. Note also that we can add the condition

$$
\mathbf{tr}((\Sigma_1 + \mu_1\mu_1^T)P) + 2\mu_1^T q + r \le 1
$$

to the problem. From these facts and the assumption that $\Sigma_1$ is positive definite, we can show that feasible solutions of $\mathrm{P}(A, b, c)$ are bounded. Therefore, the set of the optimal solutions of $\mathrm{P}(A, b, c)$ is nonempty. Let $(P^*, q^*, r^*, \tau^*)$ be one of them. Now for $(A', b', c') = (P^*, q^*, r^* - 1)$, consider the new classifier $z^T A' z + 2b'^T z + c'$. Define

$$
\begin{aligned}
p_1 &= \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x^T A x + 2b^T x + c \ge 0\}, \\
p_2 &= \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x^T A x + 2b^T x + c \le 0\},
\end{aligned} \tag{23}
$$

and also

$$
\begin{aligned}
p_3 &= \sup_{x \sim (\mu_1, \Sigma_1)} \Pr\{x^T A' x + 2b'^T x + c' \ge 0\}, \\
p_4 &= \sup_{x \sim (\mu_2, \Sigma_2)} \Pr\{x^T A' x + 2b'^T x + c' \le 0\},
\end{aligned} \tag{24}
$$

We prove the lemma by showing $p_3 \leq p_1$ and $p_4 \leq p_2$.

Note that $p_3$ is the optimal value of $P(A', b', c')$ and that $(P^*, q^*, r^*, 1)$ is a feasible solution of $P(A', b', c')$ with objective value $p_1$. Hence we have $p_3 \leq p_1$.

Next, the second constraint in $P(A, b, c)$ indicates that

$$z^T A' z + 2b'^T z + c' \geq \tau^* (z^T A z + 2b^T z + c) \quad \text{for any } z \in \Re^n.$$

Remark that from the assumption that $\alpha_q < 1$, we have $p_1 < 1$, which implies $\tau^* > 0$. Then we have

$$\{z \in \Re^n | \ z^T A z + 2b^T z + c > 0\} \subset \{z \in \Re^n | \ z^T A' z + 2b'^T z + c' > 0\}.$$

which is equivalent to

$$\{z \in \Re^n | \ z^T A' z + 2b'^T z + c' \leq 0\} \subset \{z \in \Re^n | \ z^T A z + 2b^T z + c \leq 0\}.$$

This shows $p_4 \leq p_2$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

PROOF OF THEOREM 5.1. From Theorem 5.2 and Lemma 5.2, we have $\alpha_1 = \alpha_3 \leq \alpha_2$. Since $\alpha_2 \leq \alpha_1$ from (7), we have $\alpha_1 = \alpha_2$. By the logic similar to the one appearing in the last paragraph of the proof of Theorem 5.2, we can show that any optimal solution of LMC is an optimal solution of QMC. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

## 6.   Conclusion

In this paper, we introduce QMC (Quadratic Minimax Classification problem) and CMC (Convex Minimax Classification problem) by generalizing LMC (Linear Minimax Classification problem) proposed by Lanckriet et al. [7]. We show that QMC is transformed to a parametric SDP (Semidefinite Programming problem), which is easily solved by an interior point method. In our preliminary computational experiments, we observe that all the solutions of QMC are linear. So we guess that QMC has the same solution as LMC in spite of the generality. We succeed to prove the result theoretically. We also show that the solution of CMC is obtained from the solution of LMC, although we do not have any direct method for solving CMC.

## References

[1] S. Boyd and L. Vandenberghe: *Convex Optimization* (Cambridge University Press, 2004).

[2] T. Cooke: A lower bound on the performance of the quadratic discriminant function. *Journal of Multivariate Analysis*, **89** (2004), 371–383.

[3] C.H. Hoi and M.R. Lyu: Robust face recognition using minimax probability machine. *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo* (IEEE, 2004), 1175–1178.

[4] K. Huang, H. Yang, I. King, and M.R. Lyu: Learning classifiers from imbalanced data based on biased minimax probability machine. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2004), 558–563.

[5] K. Huang, H. Yang, I. King, M.R. Lyu, and L. Chan: The minimum error minimax probability machine. *Journal of Machine Learning Research*, **5** (2004), 1253–1286.

[6] T. Kitahara, S. Mizuno, and K. Nakata: An extension of a minimax approach to multiple classification. *Journal of Operations Research Society of Japan*, **50** (2007), 123–136.

[7] G.R.G. Lanckriet, L.El Ghaoui, C. Bhattacharyya, and M.I. Jordan: A robust minimax approach to classification. *Journal of Machine Learning Research*, **3** (2002), 555–582.

[8] A. Marshall and I. Olkin: Multivariate chebyshev inequalities. *Annals of Mathematical Statistics*, **31** (1960), 1001–1014.

[9] I. Popescu: *Applications of optimization in probability, finance and revenue management* (Ph. D. thesis, Applied Mathematics Department and Operations Research Center, Massachusetts Institute of Technology, 1999).

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

Tomonari Kitahara
Department of Industrial Engineering and Management
Tokyo Institute of Technology
2-12-1 Ohokayama
Meguro Tokyo 152-8552, Japan
E-mail: `kitahara.t.ab@m.titech.ac.jp`