Optimal data fitting: a moment approach

Dimitris Bertsimas [*] Xuan Vinh Doan [†] Jean Lasserre [‡]

January 2007

Abstract

We propose a moment relaxation for two problems, the separation and covering problem with semi-algebraic sets generated by a polynomial of degree d. We show that (a) the optimal value of the relaxation finitely converges to the optimal value of the original problem, when the moment order r increases and (b) there exist probability measures such that the moment relaxation is equivalent to the original problem with $r = d$. We further provide a practical iterative algorithm that is computationally tractable for large datasets and present encouraging computational results.

1 Introduction

Data fitting problems have long been very useful in many different application areas. A well-known problem is the problem of finding the minimum-volume ellipsoid in n-dimensional space $\mathbb{R}^n$ containing all points that belong to a given finite set $S \in \mathbb{R}^n$. This minimum-volume covering ellipsoid problem is important in the area of robust statistics, data mining, and cluster analysis (see Sun and Freund [13] and references therein). Pattern separation as described in Calafiore [3] is another related problem, in which an ellipsoid that separates a set of points S_1 from another set of points S_2 needs to be found under some appropriate optimality criteria such as minimum volume or minimum distance error.

These problems have been studied for a long time. The minimum-volume covering ellipsoid problem was discussed by John [8] in 1948. Recently, this problem has been modeled as a convex optimization problem with linear matrix inequalities (LMI) and solved by interior-point methods (IPM) in Vandenberghe et al. [14], Sun and Freund [13] and Magnani et al. [11]. The problem of pattern separation via

[*]Boeing Professor of Operations Research, Sloan School of Management, co-director of the Operations Research Center, Massachusetts Institute of Technology, E40-147, Cambridge, MA 02139-4307, dbertsim@mit.edu.

[†]Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, vanxuan@mit.edu.

[‡]LAAS, 7 Avenue Du Colonel Roche, 31077 Toulouse Cédex 4, France, lasserre@laas.fr

ellipsoids was studied by Rosen [12] and Barnes [1]. Glineur [6] has proposed some methods to solve this problem with different optimality criteria via conic programming.

Although strong results were presented in Vandenberghe et al. [14], Sun and Freund [13] and Glineur [6] for these problems, they could become computationally intractable if the cardinality of datasets is large. In addition, different semi-algebraic sets other than ellipsoidal sets could be considered for these data fitting problems.

Contributions and Paper Outline

In this paper, we propose a common methodology for these data fitting problems and its extension to general semi-algebraic sets based on a moment approach. Specifically, our contributions and structure of the paper are as follows:

(1) In Section 2, we propose a moment relaxation for these data fitting problems with general semi-algebraic sets $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \theta(\boldsymbol{x}) \geq 0\}$, where $\theta \in \mathbb{R}[\boldsymbol{x}]$. We show that the optimal value of the relaxation finitely converges to the optimal value of the original problem, when the moment order r increases. The key idea of our approach is that, instead of imposing a constraint for each point in the dataset, we require that the support of any probability measure μ that is generated on the dataset is contained in Ω. Using powerful results from the theory of moments, we could replace all membership constraints to a single LMI constraint of size $\binom{n+r}{r}$.

(2) In Section 3, we show that there exist probability measures such that the moment relaxation is equivalent to the original problem with $r = \deg(\theta)$ using results on semi-infinite optimization from Ben-Tal et al. [2].

(3) In Section 4, we provide a practical iterative algorithm based on the results of Section 3 for these data fitting problems that is computationally tractable for datasets with a very large number of points. We present encouraging computational results for ellipsoidal sets and discuss the convergence of the algorithm.

2 Moment Relaxations

2.1 Problem Formulation

Consider a polynomial $\theta \in \mathbb{R}[\boldsymbol{x}]$ of degree at most d: $\theta(\boldsymbol{x}) = \sum_{\boldsymbol{\gamma} \in \mathbb{N}^n : |\boldsymbol{\gamma}| \leq d} \theta_{\boldsymbol{\gamma}} \boldsymbol{x}^{\boldsymbol{\gamma}}$. Let $\boldsymbol{\theta} = \{\theta_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \mathbb{N}^n, |\boldsymbol{\gamma}| \leq d\}$ be the coefficient vector of θ, $\boldsymbol{\theta} \in \mathbb{R}^s$, where $s = \binom{n+d}{d}$ and $\Theta \subset \mathbb{R}^s$, we obtain a family of semi-algebraic sets $\Omega_\theta = \{\boldsymbol{x} \in \mathbb{R}^n : \theta(\boldsymbol{x}) \geq 0\}$ for $\theta \in \Theta$. The problem of separating a finite dataset S_1 from another finite dataset S_2 with one of these semi-algebraic sets can be written as follows:

$$\mathcal{P}^s \triangleq \begin{bmatrix} \inf & f(\boldsymbol{\theta}) & \\ \text{s.t.} & \theta(\boldsymbol{x}) \geq 0, & \forall \boldsymbol{x} \in S_1, \\ & -\theta(\boldsymbol{x}) \geq 0, & \forall \boldsymbol{x} \in S_2, \\ & \boldsymbol{\theta} \in \Theta, & \end{bmatrix} \tag{1}$$

where f is an optimality criteria.

If we only consider one dataset S, then we can formulate the problem of covering S with the best semi-algebraic set Ω_θ with respect to optimality criterion f as follows:

$$\mathcal{P}^c \triangleq \begin{bmatrix} \inf & f(\boldsymbol{\theta}) & \\ \text{s.t.} & \theta(\boldsymbol{x}) \geq 0, & \forall \boldsymbol{x} \in S, \\ & \boldsymbol{\theta} \in \Theta. & \end{bmatrix} \tag{2}$$

If f is the volume function of ellipsoids and θ is a quadratic function that generates ellipsoidal sets, then the pattern separation via ellipsoids and minimum-volume covering ellipsoid problems are obtained respectively from these two general problems. Since the covering problem is a special case of the separation problem ($S_2 = \emptyset$), we focus on the latter problem in the following sections.

2.2 Moment Formulation

We now investigate the application of the moment approach (see Henrion [7], Lasserre [9], and the references therein) to Problem (1). Let μ^i be a probability measure generated on S_i, $i = 1, 2$,

$$\mu^i := \sum_{\boldsymbol{x} \in S_i} \mu^i_{\boldsymbol{x}} \delta_{\boldsymbol{x}}, \tag{3}$$

where $\delta_{\boldsymbol{x}}$ denotes the Dirac measure at $\boldsymbol{x}$, $\sum_{\boldsymbol{x} \in S_i} \mu^i_{\boldsymbol{x}} = 1$, and $\mu^i_{\boldsymbol{x}} \geq 0$ for all $\boldsymbol{x} \in S_i$, $i = 1, 2$. For example, the uniform probability measure μ^i generated on S_i has $\mu^i_{\boldsymbol{x}} = 1/|S_i|$ for all $\boldsymbol{x} \in S_i$.

All the moments $\boldsymbol{y}^i = \{y^i_{\boldsymbol{\alpha}}\}$ of μ^i are calculated as follows:

$$y^i_{\boldsymbol{\alpha}} = \int \boldsymbol{x}^{\boldsymbol{\alpha}} \mathrm{d}\mu^i = \sum_{\boldsymbol{x} \in S_i} \mu^i_{\boldsymbol{x}} \boldsymbol{x}^{\boldsymbol{\alpha}}, \quad \boldsymbol{\alpha} \in \mathbb{N}^n. \tag{4}$$

For any nonnegative integer r, the r-moment matrix associated with μ^i (or equivalently, with $\boldsymbol{y}^i$) $M_r(\mu^i) \equiv M_r(\boldsymbol{y}^i)$ is a matrix of size $\binom{n+r}{r}$. Its rows and columns are indexed in the canonical basis $\{\boldsymbol{x}^{\boldsymbol{\alpha}}\}$ of $\mathbb{R}[\boldsymbol{x}]$, and its elements are defined as follows:

$$M_r(\boldsymbol{y}^i)(\boldsymbol{\alpha}, \boldsymbol{\beta}) = y^i_{\boldsymbol{\alpha}+\boldsymbol{\beta}}, \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n, |\boldsymbol{\alpha}|, |\boldsymbol{\beta}| \leq r. \tag{5}$$

Similarly, given $\theta \in \mathbb{R}[\boldsymbol{x}]$, the localizing matrix $M_r(\boldsymbol{\theta y}^i)$ associated with $\boldsymbol{y}^i$ and θ is defined by

$$M_r(\boldsymbol{\theta y}^i)(\boldsymbol{\alpha},\boldsymbol{\beta}) := \sum_{\boldsymbol{\gamma}\in\mathbb{N}^n} \theta_{\boldsymbol{\gamma}} y^i_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n, |\boldsymbol{\alpha}|, |\boldsymbol{\beta}| \le r, \tag{6}$$

where $\boldsymbol{\theta} = \{\theta_{\boldsymbol{\gamma}}\}$ is the vector of coefficients of θ in the canonical basis $\{\boldsymbol{x}^{\boldsymbol{\alpha}}\}$.

If we define the matrix $M_r^{\boldsymbol{\gamma}}(\boldsymbol{y}^i)$ with elements

$$M_r^{\boldsymbol{\gamma}}(\boldsymbol{y}^i)(\boldsymbol{\alpha},\boldsymbol{\beta}) = y^i_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma} \in \mathbb{N}^n, |\boldsymbol{\alpha}|, |\boldsymbol{\beta}| \le r,$$

then the localizing matrix can be expressed as $M_r(\boldsymbol{\theta y}^i) = \sum_{\boldsymbol{\gamma}\in\mathbb{N}^n} \theta_{\boldsymbol{\gamma}} M_r^{\boldsymbol{\gamma}}(\boldsymbol{y}^i)$.

Note that for every polynomial $f \in \mathbb{R}[\boldsymbol{x}]$ of degree at most r with its vector of coefficients denoted by $\boldsymbol{f} = \{f_{\boldsymbol{\gamma}}\}$, we have:

$$\langle \boldsymbol{f}, M_r(\boldsymbol{\theta y}^i)\boldsymbol{f}\rangle = \int \theta f^2 \mathrm{d}\mu^i. \tag{7}$$

This property shows that necessarily, $M_r(\boldsymbol{\theta y}^i) \succeq \mathbf{0}$, whenever μ^i has its support contained in the level set $\{\boldsymbol{x} \in \mathbb{R}^n : \theta(\boldsymbol{x}) \ge 0\}$. Therefore, if we replace all membership constraints in the problem $\mathcal{P}^s$ by two LMI constraints $M_r(\boldsymbol{\theta y}^1) \succeq 0$ and $M_r(-\boldsymbol{\theta y}^2) \succeq 0$, we obtain a relaxation problem of $\mathcal{P}^s$:

$$\mathcal{P}_r^s(\boldsymbol{y}^1,\boldsymbol{y}^2) \triangleq \begin{bmatrix} \inf & f(\boldsymbol{\theta}) \\ \text{s.t.} & M_r(\boldsymbol{\theta y}^1) \succeq \mathbf{0} \\ & M_r(-\boldsymbol{\theta y}^2) \succeq \mathbf{0} \\ & \boldsymbol{\theta} \in \Theta. \end{bmatrix} \tag{8}$$

2.3 Convergence in Moment Order

Compared to $\mathcal{P}^s$, the data of S_1 and S_2 are aggregated into the vector $\boldsymbol{y}^1$ and $\boldsymbol{y}^2$ used in $\mathcal{P}_r^s(\boldsymbol{y}^1,\boldsymbol{y}^2)$. Both problems have exactly the same variables, but

- problem $\mathcal{P}^s$ has $|S_1| + |S_2|$ linear constraints, whereas
- problem $\mathcal{P}_r^s(\boldsymbol{y})$ has two LMI constraints $M_r(\boldsymbol{\theta y}^1) \succeq 0$ and $M_r(-\boldsymbol{\theta y}^2) \succeq 0$ with matrix size $\binom{n+r}{r}$.

If r is not too large, solving $\mathcal{P}_r^s(\boldsymbol{y})$ is preferable to solving $\mathcal{P}^s$, especially if $|S_1| + |S_2|$ is large. It is natural to ask how good this moment relaxation could be as compared to the original problem and which value of r we have to use to obtain a strong lower bound. In this section, let us assume that fixed probability measures μ^i generated on S_i, $i = 1, 2$, are selected; for example, the uniform probability measures as mentioned in the previous section.

Proposition 1 *Let $\theta \in \mathbb{R}[\boldsymbol{x}]$, and let $\mathcal{P}^s$, $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$, $r \in \mathbb{N}$ be as in (1) and (8) respectively. Then:*

$$Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z_{r+1}^s(\boldsymbol{y}^1, \boldsymbol{y}^2), \quad \text{and} \quad Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z^s, \quad \forall r \in \mathbb{N},$$

where Z^s and $Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ are optimal values of $\mathcal{P}^s$ and $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ respectively.

Proof. For every $\boldsymbol{\gamma} \in \mathbb{N}^n$, $M_r^{\boldsymbol{\gamma}}(\boldsymbol{y}^i)$ is the north-west corner square submatrix with size $\binom{n+r}{r}$ of $M_{r+1}^{\boldsymbol{\gamma}}(\boldsymbol{y}^i)$, $i = 1, 2$. This follows directly from the definition of the matrix $M_r^{\boldsymbol{\gamma}}(\boldsymbol{y}^i)$.

Since $M_r(\boldsymbol{\theta}\boldsymbol{y}^1) = \sum_{\boldsymbol{\gamma} \in \mathbb{N}^n} \theta_{\boldsymbol{\gamma}} M_r^{\boldsymbol{\gamma}}(\boldsymbol{y}^1)$ for all r, $M_r(\boldsymbol{\theta}\boldsymbol{y}^1)$ is also a north-west corner square submatrix of $M_{r+1}(\boldsymbol{\theta}\boldsymbol{y}^1)$. This implies that if $M_{r+1}(\boldsymbol{\theta}\boldsymbol{y}^1) \succeq \mathbf{0}$, then $M_r(\boldsymbol{\theta}\boldsymbol{y}^1) \succeq \mathbf{0}$. Similar arguments can be applied for $M_r(-\boldsymbol{\theta}\boldsymbol{y}^2)$ and $M_{r+1}(-\boldsymbol{\theta}\boldsymbol{y}^2)$. Thus, any feasible solution of $\mathcal{P}_{r+1}^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ is feasible in $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$. So we have:

$$Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z_{r+1}^s(\boldsymbol{y}^1, \boldsymbol{y}^2), \quad \forall r \in \mathbb{N}.$$

Similarly, any feasible solution of $\mathcal{P}^s$ is feasible in $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$. Indeed, if $\boldsymbol{\theta}$ is feasible for $\mathcal{P}^s$ then we have $\theta(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in S_1$ and $\theta(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in S_2$. Therefore, the probability measures μ^1 and μ^2 defined in (3) have their supports contained in the level set $\{\boldsymbol{x} \in \mathbb{R}^n : \theta(\boldsymbol{x}) \geq 0\}$ and $\{\boldsymbol{x} \in \mathbb{R}^n : \theta(\boldsymbol{x}) \leq 0\}$ respectively. In view of (7), we have $M_r(\boldsymbol{\theta}\boldsymbol{y}^1) \succeq \mathbf{0}$ and $M_r(-\boldsymbol{\theta}\boldsymbol{y}^2) \succeq \mathbf{0}$. This proves that $\boldsymbol{\theta}$ is feasible for $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ with any $r \in \mathbb{N}$. Thus, $Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z^s, \quad \forall r \in \mathbb{N}$. □

We next show that if μ^i is supported on the whole set S_i, $i = 1, 2$, then the optimal values $\left\{Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)\right\}$ converge to Z^s, when r increases and the convergence is finite. The statement is formally stated and proved as follows:

Theorem 1 *Let $\mathcal{P}^s$, μ^i and $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$, $r \in \mathbb{N}$ be as in (1), (3) and (8), respectively. If $\mu^i_{\boldsymbol{x}} > 0$ for all $\boldsymbol{x} \in S_i$, $i = 1, 2$, then*

$$Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \uparrow Z^s$$

and the convergence is finite.

Proof. From Proposition 1, we have $Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z^s, \quad \forall r \in \mathbb{N}$. In addition, as μ^i in (3) is finitely supported, its moment matrix $M_r(\boldsymbol{y}^i)$ defined in (5) with $\boldsymbol{y}^i$ as in (4) has finite rank. That is, there exists $r_0^i \in \mathbb{N}$ such that

$$\operatorname{rank}(M_r(\boldsymbol{y}^i)) = \operatorname{rank}(M_{r_0^i}(\boldsymbol{y})), \quad \forall r \geq r_0^i.$$

In other words, $M_r(\boldsymbol{y}^i)$ is a flat extension of $M_{r_0}(\boldsymbol{y}^i)$ for all $r \geq r_0^i$ (see Curto and Fialkow [5] for more details).

Now, let $r_0 = \max\{r_0^1, r_0^2\}$ and let $\boldsymbol{\theta}$ be an arbitrary ϵ-optimal solution of $\mathcal{P}^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2)$, $\epsilon > 0$. We have, $\boldsymbol{\theta}$ is feasible for $\mathcal{P}^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2)$ and $f(\boldsymbol{\theta}) \leq Z^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2) + \epsilon$. With $M_{r_0}(\boldsymbol{y}^i) \succeq \mathbf{0}$, using Theorem 1.6 in [5], we can deduce that μ^1 has its support contained in the level set $\{\boldsymbol{x} \in \mathbb{R}^n : \theta(\boldsymbol{x}) \geq 0\}$. Similarly, μ^2 has its support contained in the level set $\{\boldsymbol{x} \in \mathbb{R}^n : \theta(\boldsymbol{x}) \leq 0\}$. This implies that $\theta(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in S_1$ and $\theta(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in S_2$ because μ^i is supported on the whole set S_i ($\mu^i_{\boldsymbol{x}} > 0$ for all $\boldsymbol{x} \in S_i$). Thus, $\boldsymbol{\theta}$ is feasible for $\mathcal{P}^s$ and $Z^s \leq f(\boldsymbol{\theta}) \leq Z^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2) + \epsilon$. We then have

$$Z^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z^s \leq Z^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2) + \epsilon.$$

As $\epsilon > 0$ was arbitrary, $Z^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$. Since from Proposition 1 $Z^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2)$ is monotone and bounded, we obtain that $Z^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2) \uparrow Z^s$ and the convergence is finite. □

Theorem 1 provides the rationale for solving $\mathcal{P}^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2)$ instead of $\mathcal{P}^s$. However, despite the finite convergence we do not know how large the value of r_0 could be. In the next section, we will discuss how to select appropriate values r for the problem $\mathcal{P}^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2)$.

3 Convergence in Measure

In this section, we analyze how the number of points in the datasets S_1 and S_2 affects the convergence of $\mathcal{P}^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2)$. We then investigate the case when different probability measures are selected to show that there exist probability measures such that $Z^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$.

3.1 Convergence in Measure

As mentioned in the previous section, $Z^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2)$ converge to Z^s, and there exists an r_0 such that $Z^s_{r_0}(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$. We investigate next the dependence of r_0 on $|S_1|$ and $|S_2|$ and show that if the number of points in S_1 and S_2 is small enough, we will have $Z^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$ for any probability measures μ^i supported on the whole set S_i, $i = 1, 2$.

Proposition 2 *Let $\mathcal{P}^s$, μ^i, and $\mathcal{P}^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2)$ be as in (1), (3), and (8), respectively. If* $\max\{|S_1|, |S_2|\} \leq \binom{n+r}{r}$ *and $\mu^i_{\boldsymbol{x}} > 0$ for all $\boldsymbol{x} \in S_i$ for $i = 1, 2$, then $Z^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$.*

Proof. We first note that $M_r(\boldsymbol{y}^i)$ is an $\binom{n+r}{r} \times \binom{n+r}{r}$ matrix. The probability measure μ^i is supported on the whole set S_i, which has $|S_i| \leq \binom{n+r}{r}$. So we have μ^i is an $|S_i|$-atomic measure; therefore, $\operatorname{rank}(M_r(\boldsymbol{y}^i)) = |S_i|$ (see Curto and Fialkow [4] for more details).

Then according to Laurent [10], there exists $|S_i|$ interpolation polynomials $f_j \in \mathbb{R}[\boldsymbol{x}]$ of degree at most r, $j = 1, \ldots, |S_i|$, such that

$$f_j(\boldsymbol{x}(k)) = \begin{cases} 0, & j \neq k, \\ 1, & j = k, \end{cases} \qquad \forall j, k = 1, \ldots, |S_i|.$$

Now, let $\boldsymbol{\theta}$ be an arbitrary ϵ-optimal solution of $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$, $\epsilon > 0$. We have, $\boldsymbol{\theta}$ is feasible for $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$, which means $M_r(\boldsymbol{\theta}\boldsymbol{y}^1) \succeq \mathbf{0}$, $M_r(-\boldsymbol{\theta}\boldsymbol{y}^2) \succeq \mathbf{0}$ and $f(\boldsymbol{\theta}) \leq Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) + \epsilon$. For every $j = 1, \ldots, |S_1|$, we have:

$$\langle \boldsymbol{f}_j, M_r(\boldsymbol{\theta}\boldsymbol{y}^1)\boldsymbol{f}_j \rangle \geq 0,$$

where $\boldsymbol{f}_j$ is the vector of coefficients of the polynomial f_j.

Eq. (7) implies that

$$\int \theta f_j^2 \mathrm{d}\mu^1 \geq 0 \Leftrightarrow \mu^1_{\boldsymbol{x}(j)}\theta(\boldsymbol{x}_j) \geq 0.$$

Since $\mu^1_{\boldsymbol{x}} > 0$ for all $\boldsymbol{x} \in S_1$, we obtain $\theta(\boldsymbol{x}(j)) \geq 0$ for all $j = 1, \ldots, |S|_1$. Similarly, we also obtain $\theta(\boldsymbol{x}(j)) \leq 0$ for all $j = 1, \ldots, |S|_2$. Thus, $\boldsymbol{\theta}$ is feasible for $\mathcal{P}^s$ and $Z^s \leq f(\boldsymbol{\theta}) \leq Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) + \epsilon$. Combining with results from Proposition 1, we have

$$Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z^s \leq Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) + \epsilon.$$

As $\epsilon > 0$ was arbitrarily chosen, we obtain $Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$. □

Proposition 2 implies that if we select $r_0 = \min\left\{r \in \mathbb{N} : \binom{n+r}{r} \geq \max\{|S_1|, |S_2|\}\right\}$, then we have $Z_{r_0}^s(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$ for any probability measures μ^i supported on the whole set S_i, $i = 1, 2$. Although the result is interesting, it is not very useful for practical algorithms. The problem $\mathcal{P}_{r_0}^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ has only two LMI constraints but its matrix size is at least $\max\{|S_1|, |S_2|\}$, which could be very large. It means the problem $\mathcal{P}_{r_0}^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ is still computationally difficult to solve, when $|S_1|$ or $|S_2|$ is large. In the next section, we will use results from this proposition to show that we can find appropriate probability measures μ^i such that $Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) = Z^s$ for r as small as d, which is the degree of the polynomial θ.

3.2 Optimal Measure

The probability measure μ^i is defined in (3) as $\mu^i = \sum_{\boldsymbol{x}\in S_i} \mu_{\boldsymbol{x}}\delta_{\boldsymbol{x}}$ with $\sum_{\boldsymbol{x}\in S_i} \mu_{\boldsymbol{x}} = 1$ and $\mu_{\boldsymbol{x}} \geq 0$ for all $\boldsymbol{x} \in S_i$. Let $\boldsymbol{\mu}^i = \left(\mu^i_{\boldsymbol{x}_1}, \ldots, \mu^i_{\boldsymbol{x}_{|S|_i}}\right)'$, we have, $\boldsymbol{\mu}^i \in M_{|S_i|}$, where $M_{|S_i|} = \left\{\boldsymbol{x} \in \mathbb{R}_+^{|S_i|} : \sum_{i=1}^{|S_i|} x_i = 1\right\}$. Each probability measure μ^i can then be represented equivalently by a vector $\boldsymbol{\mu}^i \in M_{|S_i|}$. Thus, the optimal value of the problem $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ can also be expressed as $Z_r^s(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2) \equiv Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$.

Clearly, we can form infinitely many moment relaxations from different probability measures generated on S_i as above. The question is then which pair of probability measures yields the best relaxation. Let us consider the following problem

$$\mathcal{P}_r^s \triangleq \left[\sup_{\boldsymbol{\mu}^i \in M_{|S_i|}} Z_r^s(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2) \right]. \tag{9}$$

We then immediately have the following result

Proposition 3 *Let $\mathcal{P}^s$ and $\mathcal{P}_r^s$ be as in (1) and (9) respectively. Then $Z_r^s \leq Z^s$, where Z_r^s is the optimal value of $\mathcal{P}_r^s$.*

Proof. From Proposition 1, we have $Z_r^s(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2) \equiv Z_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2) \leq Z^s$ for all $\boldsymbol{\mu}^i \in M_{|S_i|}$, $i = 1, 2$. We have:

$$Z_r^s = \sup_{\boldsymbol{\mu}^i \in M_{|S_i|}} Z_r^s(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2).$$

Thus, $\sup_{\boldsymbol{\mu}^i \in M_{|S_i|}} Z_r^s(\boldsymbol{\mu}^1, \boldsymbol{\mu}^2) \leq \sup_{\boldsymbol{\mu}^i \in M_{|S_i|}} Z^s = Z^s$, and thus $Z_r^s \leq Z^s$. □

We are interested in finding the minimum value of the moment order r that makes the above inequality become equality. We observe that the optimal solution of $\mathcal{P}^s$ depends only on some small subsets of S_1 and S_2 given the convexity of f and Θ. The following theorem is proved by Ben-Tal et al. [2]:

Theorem 2 (Ben-Tal et al. [2]) *Consider the problem*

$$\mathcal{P} \triangleq \left[\begin{array}{lll} \inf & f(\boldsymbol{\theta}) & \\ \text{s.t.} & g_k(\boldsymbol{\theta}, \boldsymbol{x}) \leq 0, \quad \boldsymbol{x} \in S_k, \quad k = 1, \ldots, m, \\ & \boldsymbol{\theta} \in \Theta, & \end{array} \right]$$

and assume that

(A1) Θ is convex with non empty interior,

(A2) f is continuous and convex on Θ,

(A3) g_k is continuous in $\boldsymbol{x}$,

(A4) g_k is continuous and convex in $\boldsymbol{\theta}$ on Θ and $\{\boldsymbol{\theta} : g_k(\boldsymbol{\theta}, \boldsymbol{x}) < 0\}$ is open $\forall k$ and $\forall \boldsymbol{x} \in S_k$,

(A5) (Slater condition) $\{\boldsymbol{\theta} \in \mathbb{R}^s : g_k(\boldsymbol{\theta}, \boldsymbol{x}) < 0, \boldsymbol{x} \in S_k, k = 1, \ldots, m\}$ is nonempty.

Let $\boldsymbol{\theta}^*$ be a feasible solution of $\mathcal{P}$ and $S_k(\boldsymbol{\theta}^*) = \{\boldsymbol{x} \in S_k : g_k(\boldsymbol{\theta}^*, \boldsymbol{x}) = 0\}$, $K^* = \{k : S_k(\boldsymbol{\theta}^*) \neq \emptyset\}$, then $\boldsymbol{\theta}^*$ is an optimal solution of $\mathcal{P}^s$ if and only if there is a set $S^* \subset \cup_{k \in K^*} S_k$ with at most s elements such that $\boldsymbol{\theta}^*$ is the optimal solution of the problem:

$$\mathcal{P}^* \triangleq \left[\begin{array}{ll} \inf & f(\boldsymbol{\theta}) \\ \text{s.t.} & g_k(\boldsymbol{\theta}, \boldsymbol{x}) \leq 0, \quad \boldsymbol{x} \in S_k \cap S^*, \quad k \in K^*, \\ & \boldsymbol{\theta} \in \Theta. \end{array} \right]$$

Using the previous theorem, we now state and prove the following result.

Theorem 3 *Let $\theta \in \mathbb{R}[\boldsymbol{x}]$ and let $\mathcal{P}^s$, $\mathcal{P}^s_r$ be defined as in (1), and (9) and assume that Θ is convex, f is convex on Θ and Slater condition is satisfied. If $\mathcal{P}^s$ is solvable then*

$$Z^s_r = Z^s, \quad \forall r \geq d,$$

where d is the degree of θ.

Proof. From Proposition 3, we have $Z^s_r \leq Z^s$ for all $r \in \mathbb{N}$.

$\theta(\boldsymbol{x})$ is a polynomial in $\boldsymbol{x}$ and linear in the coefficient vector $\boldsymbol{\theta}$. In addition, $\mathcal{P}^s$ is solvable; therefore, we can then apply the result from the previous theorem. Consider the set S^* that defines the reduced problem $\mathcal{P}^*$, we have: $|S^*| \leq s = \binom{n+d}{d}$. Choose probability measures μ^i_0 (with the moment vector $\boldsymbol{y}^i_0$) supported exactly on the whole set $S^* \cap S_i$, that is, for all $\boldsymbol{x} \in S_i$, $\mu^i_{\boldsymbol{x}} > 0$ if and only if $\boldsymbol{x} \in S^*$. Clearly, $\boldsymbol{\mu}^i_0 \in M_{|S_i|}$, thus $Z^s_r(\boldsymbol{\mu}^1_0, \boldsymbol{\mu}^2_0) \leq Z^s_r$.

Let $\mathcal{P}^s(S^*)$ denote the problem $\mathcal{P}^s$ constructed on $S^* \cap S_i$ instead of S_i and $Z^s(S^*)$ be its optimal value. The probability measure μ^i_0 is supported on the whole set $S^* \cap S_i$, thus $\mathcal{P}^s_r(\boldsymbol{y}^1_0, \boldsymbol{y}^2_0)$ is also a moment relaxation problem of $\mathcal{P}^s(S^*)$. We have, $|S^* \cap S_i| \leq |S^*| \leq \binom{n+r}{r}$, $i = 1, 2$, for all $r \geq d$. Applying the result from Proposition 2, we then have $Z^s_r(\boldsymbol{\mu}^1_0, \boldsymbol{\mu}^2_0) = Z^s(S^*)$.

The optimal solution of $\mathcal{P}^s(S^*)$ is also the optimal solution for $\mathcal{P}^s$, which means $Z^s(S^*) = Z^s$. From these inequalities and equalities, we have: $Z^s = Z^s(S^*) = Z^s_r(\boldsymbol{\mu}^1_0, \boldsymbol{\mu}^2_0) \leq Z^s_r \leq Z^s$ for all $r \geq d$. □

The above theorem shows that with r as small as d, the moment relaxation $\mathcal{P}^s_r(\boldsymbol{y}^1, \boldsymbol{y}^2)$ is equivalent to the original problem $\mathcal{P}^s$, given that the appropriate probability measures are used. In the next section, we propose a practical algorithm to find a sequence of probability measures such that the original problem can be solved by the corresponding sequence of relaxations.

4 Practical Algorithm

4.1 Algorithm

Theorem 3 indicates that if we solve Problem $\mathcal{P}_r^s$ for $r \geq d$, then we solve the original problem $\mathcal{P}^s$, assuming that $\mathcal{P}^s$ is solvable. The key question is how to select the optimal probability measures for problem $\mathcal{P}_r^s$. The proof of Theorem 3 suggests that in order to find the optimal probability measures, we need to find the set of points S^* that defines the optimal solution of $\mathcal{P}^s$. We propose an iterative algorithm, in each iteration of which, we solve Problem $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ with different μ^1 and μ^2 until we find the optimal probability measures.

The main algorithm is described as follows:

Main Algorithm

1. Initialization. Set $k \leftarrow 0$, $S^k \leftarrow S_1 \cup S_2$, $r \leftarrow d$.

2. Create μ_k^i uniformly over $S^k \cap S_i$. Solve $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$. Obtain optimal solution θ_k.

3. Form set of outside points $O^k = \{\boldsymbol{x} \in S_1 : \theta_k(\boldsymbol{x}) < 0\} \cup \{\boldsymbol{x} \in S_2 : \theta_k(\boldsymbol{x}) > 0\}$. If $O^k = \emptyset$, STOP. Return $\theta = \theta_k$.

4. Update $k \leftarrow k+1$, $S^k \leftarrow \{\boldsymbol{x} \in S_1 : \theta_k(\boldsymbol{x}) \leq 0\} \cup \{\boldsymbol{x} \in S_2 : \theta_k(\boldsymbol{x}) \geq 0\}$. Go to step 1.

The update rule for supporting sets is based on the fact that points outside the current optimal set obtained from the moment relaxation problem are likely to be among the points that define the optimal separation (or covering) set. This is also the reason why S^0 is selected as $S_1 \cup S_2$, which helps to separate critical and non-critical points right after the first iteration. After a supporting set are created, all points in the set are to be equally considered; therefore, uniform probability measures are used to form the moment relaxation problem in each iteration.

When the algorithm terminates, we indeed obtain an optimal solution to our original problem $\mathcal{P}^s$. Let us assume the algorithm terminates when $k = K$, then the set of outside points $O^K = S_1^K \cup S_2^K$, where $S_1^K = \{\boldsymbol{x} \in S_1 : \theta_K(\boldsymbol{x}) < 0\}$ and $S_2^K = \{\boldsymbol{x} \in S_2 : \theta_K(\boldsymbol{x}) > 0\}$, is empty. Thus the optimal solution of Problem $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ with the uniform distribution μ^{iK} over $S_i^K \subset S_i$ is a feasible solution to Problem $\mathcal{P}^s$. We then have $Z_r^s(\boldsymbol{\mu}^{1K}, \boldsymbol{\mu}^{2K}) \geq Z^s$, where $Z_r^s(\boldsymbol{\mu}^{1K}, \boldsymbol{\mu}^{2K})$ is the optimal value of Problem $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$. On the other hand, Problem $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ is a relaxation of Problem $\mathcal{P}^s(S_1^K, S_2^K)$, the problem $\mathcal{P}^s$ constructed over S_i^K instead of S_i, $i = 1, 2$. Therefore, $Z_r^s(\boldsymbol{\mu}^{1K}, \boldsymbol{\mu}^{2K}) \leq Z^s(S_1^K, S_2^K)$, where $Z^s(S_1^K, S_2^K)$ is the optimal value of Problem $\mathcal{P}^s(S_1^K, S_2^K)$. In addition, $S_i^K \subset S_i$, thus $Z^s(S_1^K, S_2^K) \leq Z^s$. Combining these inequalities, we obtain $Z_r^s(\boldsymbol{\mu}^{1K}, \boldsymbol{\mu}^{2K}) = Z^s$ or the optimal solution to Problem $\mathcal{P}_r^s(\boldsymbol{y}^1, \boldsymbol{y}^2)$ is an optimal solution to Problem $\mathcal{P}^s$.

In the following sections, we will consider the minimum-volume covering ellipsoids problem and the separation problem via ellipsoids. Some computational results with this iterative algorithm for these two problems will be reported.

4.2 Minimum-Volume Covering Ellipsoid Problem

4.2.1 Problem Formulations

The minimum-volume covering ellipsoid problem involves only one dataset. Let $S \in \mathbb{R}^n$ be a finite set of points, $S = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s\}$, where $s = |S|$. We assume that the affine hull of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$ spans $\mathbb{R}^n$, which will guarantee any ellipsoids that cover all the points in S have positive volume.

The ellipsoid $\Omega \in \mathbb{R}^n$ to be determined can be written

$$\Omega := \left\{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x} - \boldsymbol{d})'\boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{d}) \leq 1\right\},$$

where $\boldsymbol{Q} \succ 0$, $\boldsymbol{Q} = \boldsymbol{Q}'$. The volume of Ω is proportional to $\det \boldsymbol{Q}^{-1/2}$; therefore, the minimum-volume covering ellipsoid problem can be formulated as a maximum determinant problem (see Vandenberghe et al. [14] for more details) as follows:

$$\mathcal{P} \triangleq \left[\begin{array}{rll} \inf_{\boldsymbol{Q},\boldsymbol{d}} & \det \boldsymbol{Q}^{-1/2} & \\ \text{s.t.} & (\boldsymbol{x} - \boldsymbol{d})'\boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{d}) \leq 1, & \forall \boldsymbol{x} \in S, \\ & \boldsymbol{Q} = \boldsymbol{Q}' \succ \boldsymbol{0}. & \end{array}\right] \tag{10}$$

Let $\boldsymbol{A} = \boldsymbol{Q}^{1/2}$ and $\boldsymbol{a} = \boldsymbol{Q}^{1/2}\boldsymbol{d}$, then $\mathcal{P}$ is equivalent to a convex optimization problem with $|S| + 1$ LMI constraints in the unknown variables $\boldsymbol{A}$ and $\boldsymbol{a}$. Indeed, each constraint $(\boldsymbol{x} - \boldsymbol{d})'\boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{d}) \leq 1$ for $\boldsymbol{x} \in S$ can be rewritten as follows:

$$\left[\begin{array}{cc} \boldsymbol{I} & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{a} \\ (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{a})' & 1 \end{array}\right] \succeq \boldsymbol{0}, \quad \boldsymbol{x} \in S,$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix.

Instead of using $\boldsymbol{A}$ and $\boldsymbol{a}$, let consider $\boldsymbol{b} = 2\boldsymbol{Q}\boldsymbol{d}$ and $\boldsymbol{c} = 1 - \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b}$, then the ellipsoid Ω can be written:

$$\Omega := \left\{\boldsymbol{x} \in \mathbb{R}^n : -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0\right\}.$$

The minimum-volume covering ellipsoid problem is formulated as follows:

$$\begin{array}{rl} \inf_{\boldsymbol{Q},\boldsymbol{b},c} & \det \boldsymbol{Q}^{-1/2} \\ \text{s.t.} & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0, \quad \forall \boldsymbol{x} \in S, \\ & \boldsymbol{c} = 1 - \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b}, \\ & \boldsymbol{Q} = \boldsymbol{Q}' \succ \mathbf{0}. \end{array} \tag{11}$$

Consider the relaxation problem

$$\begin{array}{rl} \inf_{\boldsymbol{Q},\boldsymbol{b},c} & \det \boldsymbol{Q}^{-1/2} \\ \text{s.t.} & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0, \quad \forall \boldsymbol{x} \in S, \\ & \boldsymbol{c} \leq 1 - \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b}, \\ & \boldsymbol{Q} = \boldsymbol{Q}' \succ \mathbf{0}. \end{array} \tag{12}$$

Lemma 1 *Any optimal solution (Q^*, b^*, c^*) of Problem (12) is an optimal solution of Problem (11).*

Proof. Since Problem (12) is an relaxation of Problem (11), we just need to prove that (Q^*, b^*, c^*) is a feasible solution of Problem (11).

Suppose there exists an optimal solution (Q, b, c) of Problem (12) that satisfies the inequality $\gamma = c + \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b} < 1$. We have:

$$-\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0 \Leftrightarrow c + \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b} \geq (\boldsymbol{Q}^{1/2}\boldsymbol{x} - \frac{1}{2}\boldsymbol{Q}^{-1/2}\boldsymbol{b})'(\boldsymbol{Q}^{1/2}\boldsymbol{x} - \frac{1}{2}\boldsymbol{Q}^{-1/2}\boldsymbol{b}) \geq 0.$$

If we assume that $|S| > 1$ then we have $\gamma = c + \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b} > 0$. Thus $0 < \gamma < 1$.

Let consider the solution $(\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{b}}, \tilde{c})$ that satisfies $\boldsymbol{Q} = \gamma\tilde{\boldsymbol{Q}}$, $\boldsymbol{b} = \gamma\tilde{\boldsymbol{b}}$, and $\tilde{c} = 1 - \frac{1}{4}\tilde{\boldsymbol{b}}'\tilde{\boldsymbol{Q}}^{-1}\tilde{\boldsymbol{b}}$, we have:

$$(\boldsymbol{Q}^{1/2}\boldsymbol{x} - \frac{1}{2}\boldsymbol{Q}^{-1/2}\boldsymbol{b})'(\boldsymbol{Q}^{1/2}\boldsymbol{x} - \frac{1}{2}\boldsymbol{Q}^{-1/2}\boldsymbol{b}) = \gamma(\tilde{\boldsymbol{Q}}^{1/2}\boldsymbol{x} - \frac{1}{2}\tilde{\boldsymbol{Q}}^{-1/2}\tilde{\boldsymbol{b}})'(\tilde{\boldsymbol{Q}}^{1/2}\boldsymbol{x} - \frac{1}{2}\tilde{\boldsymbol{Q}}^{-1/2}\tilde{\boldsymbol{b}})$$

Thus

$$-\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0 \Leftrightarrow (\tilde{\boldsymbol{Q}}^{1/2}\boldsymbol{x} - \frac{1}{2}\tilde{\boldsymbol{Q}}^{-1/2}\tilde{\boldsymbol{b}})'(\tilde{\boldsymbol{Q}}^{1/2}\boldsymbol{x} - \frac{1}{2}\tilde{\boldsymbol{Q}}^{-1/2}\tilde{\boldsymbol{b}}) \leq 1,$$

or we have $-\boldsymbol{x}'\tilde{\boldsymbol{Q}}\boldsymbol{x} + \tilde{\boldsymbol{b}}'\boldsymbol{x} + \tilde{c} \geq 0$ for all $\boldsymbol{x} \in S$. Therefore, the solution $(\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{b}}, \tilde{c})$ is a feasible for Problem (12). However, we have:

$$\boldsymbol{Q} = \gamma\tilde{\boldsymbol{Q}} \Rightarrow \det \tilde{\boldsymbol{Q}}^{-1/2} = \gamma^{n/2} \det \boldsymbol{Q}^{-1/2} < \det \boldsymbol{Q}^{-1/2}.$$

This contradicts the fact that $(\boldsymbol{Q}, \boldsymbol{b}, c)$ is an optimal solution of Problem (12). Thus we must have $\boldsymbol{c} = 1 - \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b}$ or $(\boldsymbol{Q}, \boldsymbol{b}, c)$ is a feasible (optimal) solution of Problem (11). □

Using Lemma 1 and the following fact:

$$\begin{cases} \frac{1}{4}\boldsymbol{b}'\boldsymbol{Q}^{-1}\boldsymbol{b} \leq 1 - c \\ \boldsymbol{Q} = \boldsymbol{Q} \succeq \boldsymbol{0} \end{cases} \Leftrightarrow \begin{bmatrix} \boldsymbol{Q} & \frac{1}{2}\boldsymbol{b} \\ \frac{1}{2}\boldsymbol{b}' & 1 - c \end{bmatrix} \succeq \boldsymbol{0},$$

we can then formulate the minimum-volume covering problem as the following maximum determinant problem with $|S|$ linear constraints:

$$\mathcal{P} \triangleq \begin{bmatrix} \inf_{\boldsymbol{Q},\boldsymbol{b},c} & \log\det \boldsymbol{Q}^{-1} \\ \text{s.t.} & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0, \quad \forall \boldsymbol{x} \in S, \\ & \begin{bmatrix} \boldsymbol{Q} & \frac{1}{2}\boldsymbol{b} \\ \frac{1}{2}\boldsymbol{b}' & 1 - c \end{bmatrix} \succeq \boldsymbol{0}. \end{bmatrix} \tag{13}$$

Clearly, with this formulation, the minimum-covering ellipsoid problem is one of the covering problems $\mathcal{P}^c$ as shown in (2).

4.2.2 Computational Results

We have implemented the algorithm presented in Section 4.1 with $r = 2$ for the minimum-volume covering ellipsoid problem (we just need to set $S_2 = \emptyset$). Datasets are generated using several independent normal distributions to represent data from one or more clusters. The data are then affinely transformed so that the geometric mean is the origin and all data points are in the unit ball. This affine transformation is done to make sure that data samples have the same magnitude. Computation is done in Matlab 7.1.0.183 (R14, SP3) with general-purpose YALMIP 3 interface and SDPT3 3.3 solver on a Pentium III 1GHz with 256MB RAM. Clearly, this algorithm can be implemented with SeDuMi or *maxdet* solver in particular for this determinant maximization problem.

The test cases show that the algorithm works well with data in two or three dimensions. Figure 1 shows the minimum-volume covering ellipsoids for a 1000-point dataset on the plane. When $n = 3$, we have run the algorithm for datasets with up to $100,000$ points. The number of iterations we need is about 6 and it decreases when we decrease the number of points to be covered. We also have results for datasets with $10,000$ points when $n = 10$. However, time to prepare moment matrices increases significantly in terms of dimension. We need to prepare $O(n^2)$ square matrices of size $O(n^2)$ as data input for the relaxation problem if $r = 2$. If the probability measure is supported on m points, then the total computational time to prepare all necessary moment matrices is proportional to $O(n^7 m)$. Clearly, this algorithm is more suitable for datasets in low dimensions with a large number of points. The computational time could be reduced significantly if we implement additional heuristics to find a good

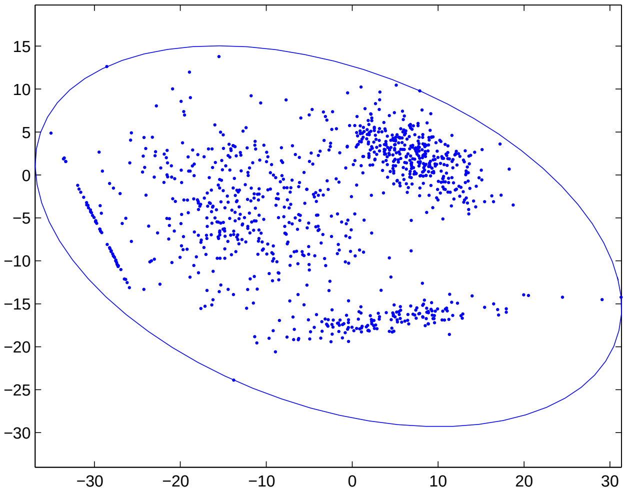

Figure 1: Minimum-volume covering ellipsoid for a 1000-point dataset.

initial subset instead of the whole set. A problem-specific SDP code that exploits the data structure of the relaxation problem could be useful for datasets in higher dimensions.

4.3 Separation Problem via Ellipsoids

4.3.1 Problem Formulation

The separation problem via ellipsoids with two datasets S_1 and S_2 is to find an ellipsoid that contains one set, for example, S_1, but not the other, which is S_2 in this case. If we represent the ellipsoid as the set $\Omega := \left\{ \boldsymbol{x} \in \mathbb{R}^n : -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0 \right\}$ with $\boldsymbol{Q} = -\boldsymbol{Q} \succ \boldsymbol{0}$, then similar to the minimum-volume covering ellipsoid problem, we can formulate the separation problem as follows:

$$\mathcal{P} \triangleq \left[\begin{array}{lll} \inf_{\boldsymbol{Q},\boldsymbol{b},c} & \log\det \boldsymbol{Q}^{-1} & \\ \quad\text{s.t.} & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0, & \forall \boldsymbol{x} \in S_1, \\ & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \leq 0, & \forall \boldsymbol{x} \in S_2, \\ & \begin{bmatrix} \boldsymbol{Q} & \frac{1}{2}\boldsymbol{b} \\ \frac{1}{2}\boldsymbol{b}' & 1-c \end{bmatrix} \succeq \boldsymbol{0}. & \end{array} \right] \tag{14}$$

4.3.2 Computational Results

Similar to the minimum-volume ellipsoid problem, the algorithm for this separation problem can be implemented with $r = 2$. With YALMIP interface and SDPT3 solver, the logdet objective function is converted to geometric mean function, which is $-(\det \boldsymbol{Q})^{1/n}$. If the problem is feasible, the optimal solution will have $\boldsymbol{Q} \succ \boldsymbol{0}$, which mean the objective value is strictly negative. This can be considered as a sufficient condition to determine whether the problem is infeasible. In each iteration of the algorithm, if the optimal value is zero ($\boldsymbol{Q} = \boldsymbol{0}$, $\boldsymbol{b} = \boldsymbol{0}$, and $c = 0$ is a feasible solution for the subproblem solved in each iteration), then we can stop and conclude that the problem is infeasible. Existence of the critical subset that determines the problem infeasibility can be proved using the same arguments as in the proof of Theorem 3 for the feasibility problem:

$$\mathcal{P} \triangleq \left[\begin{array}{lll} \inf_{\boldsymbol{Q},\boldsymbol{b},c,d} & d & \\ \quad\text{s.t.} & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \geq 0, & \forall \boldsymbol{x} \in S_1, \\ & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \leq d, & \forall \boldsymbol{x} \in S_2, \\ & \begin{bmatrix} \boldsymbol{Q} & \frac{1}{2}\boldsymbol{b} \\ \frac{1}{2}\boldsymbol{b}' & 1-c \end{bmatrix} \succeq \boldsymbol{0}. & \end{array} \right] \tag{15}$$

In order to test the algorithm, we generate datasets S_1 and S_2 as for the minimum-volume covering ellipsoids problem. In most of the cases, if we run the algorithm for S_1 and S_2, we will get the infeasibility result. In order to generate separable datasets, we run the minimum-volume covering ellipsoid algorithm for S_1 and generate the separable set S_2' from S_2 by selecting all points that are outside the ellipsoid. We also try to include some points that are inside the ellipsoid to test the cases when S_1 and S_2' are separable by a different ellipsoid rather than the minimum-volume ellipsoid that covers S_1.

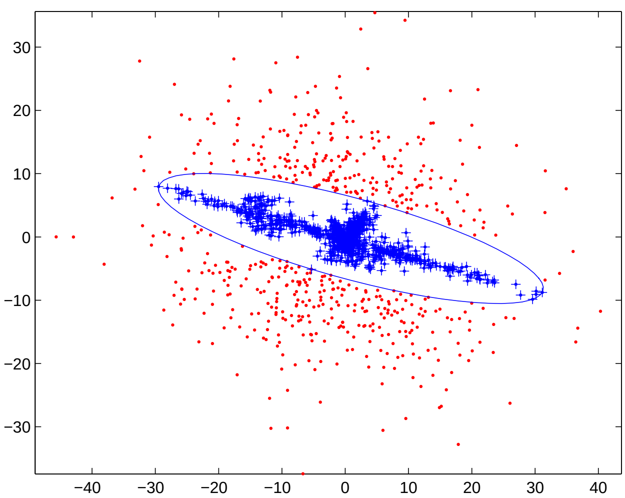

Figure 2: Separating ellipsoid is the same as the minimum-volume covering ellipsoid.

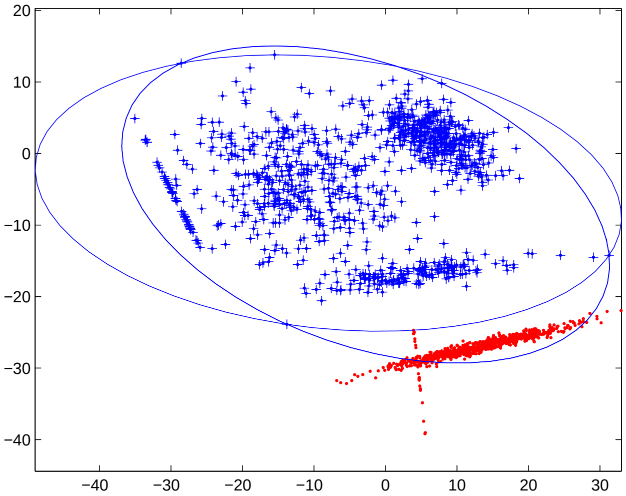

Figure 3: Separating ellipsoid is different from the minimum-volume covering ellipsoid.

The test cases show that the algorithm can detect problem infeasibility and in the separable case, finding the ellipsoid that separates two datasets. Figure 2 shows the separation of two datasets on the plane with 1000 points by minimum-volume ellipsoid while Figure 3 represents the case when a different ellipsoid is needed to separate two particular sets. We also run the algorithm for datasets with $n = 3$ and $n = 10$. Similar remarks can be made with respect to data preparation and other algorithmic issues as in Section 4.2.2. In general, the algorithm is suitable for datasets in low dimensions with a large number of points.

References

[1] E. R. Barnes. An algorithm for separating patterns by ellipsoids. *IBM Journal of Research and Development*, 26(6):759–764, 1982.

[2] A. Ben-Tal, E.E. Rosinger, and A. Ben-Israel. A Helly-type theorem and semi-infinite programming. In C.V. Coffman and G.J. Fix, editors, *Constructive Approaches to Mathematical Models*, pages 127–135. Academic Press, New York, 1979.

[3] G. Calafiore. Approximation of n-dimensional data using spherical and ellipsoidal primitives. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 32(2):269–278, 2002.

[4] R. E. Curto and L. A. Fialkow. Solution of the truncated complex moment problem for flat data. In *Memoirs of the American Mathematical Society*, volume 19. American Mathematical Society, Providence, RI, 1996.

[5] R. E. Curto and L. A. Fialkow. The truncated complex K-moment problem. *Transactions of the American Mathematical Society*, 352(6):2825–2855, 2000.

[6] F. Glineur. Pattern separation via ellipsoids and conic programming. In *Mémoire de DEA*. Faculté Polytechnique de Mons, Belgium, September 1998.

[7] D. Henrion and J. B. Lasserre. Solving nonconvex optimization problems. *IEEE Control System Magazine*, 24:72–83, 2004.

[8] F. John. Extreme problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday*, pages 187–204. Wiley Interscience, New York, 1948.

[9] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[10] M. Laurent. Revisiting two theorems of Curto and Fialkow on moment matrices. *Proceedings of the American Mathematical Society*, 133(10):2965–2976, 2005.

[11] A. Magnani, S. Lall, and S. Boyd. Tractable fitting with convex polynomials via sum-of-squares. In *Proceedings of the 44th IEEE Conference on Decision and Control*, Seville, Spain, December 2005.

[12] J. B. Rosen. Pattern separation by convex programming. *Journal of Mathematical Analysis and Applications*, 10:123–134, 1965.

[13] P. Sun and R. M. Freund. Computation of minimum-volume covering ellipsoids. *Operations Research*, 52(5):690–706, 2004.

[14] L. Vandenberghe, S. Boyd, and S. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.