

2-2016

The Log-Exponential Smoothing Technique and Nesterov's Accelerated Gradient Method for Generalized Sylvester Problems

N. T. An

Thua Thien Hue College of Education

Daniel J. Giles

Portland State University, dangiles@pdx.edu

Nguyen Mau Nam

Portland State University, mau.nam.nguyen@pdx.edu

R. Blake Rector

Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/mth_fac

 Part of the [Algebraic Geometry Commons](#)

Citation Details

An, N. T.; Giles, Daniel J.; Nam, Nguyen Mau; and Rector, R. Blake, "The Log-Exponential Smoothing Technique and Nesterov's Accelerated Gradient Method for Generalized Sylvester Problems" (2016). *Mathematics and Statistics Faculty Publications and Presentations*. 155.

https://pdxscholar.library.pdx.edu/mth_fac/155

This Pre-Print is brought to you for free and open access. It has been accepted for inclusion in Mathematics and Statistics Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

THE LOG-EXPONENTIAL SMOOTHING TECHNIQUE AND NESTEROV'S ACCELERATED GRADIENT METHOD FOR GENERALIZED SYLVESTER PROBLEMS

N.T. An¹, D. Giles², N.M. Nam³, R. B. Rector⁴.

Abstract: The Sylvester smallest enclosing circle problem involves finding the smallest circle that encloses a finite number of points in the plane. We consider generalized versions of the Sylvester problem in which the points are replaced by sets. Based on the log-exponential smoothing technique and Nesterov's accelerated gradient method, we present an effective numerical algorithm for solving these problems.

Key words. log-exponential smoothing; minimization majorization algorithm; Nesterov's accelerated gradient method; generalized Sylvester problem.

AMS subject classifications. 49J52, 49J53, 90C31.

1 Introduction

The *smallest enclosing circle problem* can be stated as follows: Given a finite set of points in the plane, find the circle of smallest radius that encloses all of the points. This problem was introduced in the 19th century by the English mathematician James Joseph Sylvester (1814–1897) [24]. It is both a facility location problem and a major problem in computational geometry. Over a century later, the smallest enclosing circle problem remains very active due to its important applications to clustering, nearest neighbor search, data classification, facility location, collision detection, computer graphics, and military operations. The problem has been widely treated in the literature from both theoretical and numerical standpoints; see [1, 4, 6, 7, 9, 21, 23, 25, 27, 28, 31] and the references therein.

The authors' recent research focuses on generalized Sylvester problems in which the given points are replaced by sets. Besides the intrinsic mathematical motivation, this question appears in more complicated models of facility location in which the sizes of the locations are not negligible, as in bilevel transportation problems. The main goal of this paper is to

¹Thua Thien Hue College of Education, 123 Nguyen Hue, Hue City, Vietnam (email: thuan2784@gmail.com).

²Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, PO Box 751, Portland, OR 97207, United States (email: dangiles@pdx.edu). The research of Daniel Giles was partially supported by the USA National Science Foundation under grant DMS-1411817.

³Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, PO Box 751, Portland, OR 97207, United States (email: mau.nam.nguyen@pdx.edu). The research of Nguyen Mau Nam was partially supported by the USA National Science Foundation under grant DMS-1411817 and the Simons Foundation under grant #208785.

⁴Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, PO Box 751, Portland, OR 97207, United States (email: r.b.rector@pdx.edu)

develop an effective numerical algorithm for solving the *smallest intersecting ball problem*. This problem asks for the smallest ball that intersects a finite number of convex target sets in \mathbb{R}^n . Note that when the target sets given in the problem are singletons, the smallest intersecting ball problem reduces to the classical Sylvester problem.

The smallest intersecting ball problem can be solved by minimizing a nonsmooth optimization problem in which the objective function is the maximum of the distances to the target sets. The nondifferentiability of this objective function makes it difficult to develop effective numerical algorithms for solving the problem. A natural approach is to approximate the nonsmooth objective function by a smooth function that is favorable for applying available smooth optimization schemes. Based on the log-exponential smoothing technique and Nesterov's accelerated gradient method, we present an effective numerical algorithm for solving this problem.

Our paper is organized as follows. Section 2 contains tools of convex optimization used throughout the paper. In Section 3, we focus the analysis of the log-exponential smoothing technique applied to the smallest intersecting ball problem. Section 4 is devoted to developing an effective algorithm based on the minimization majorization algorithm and Nesterov's accelerated gradient method to solve the problem. We also analyze the convergence of the algorithm. Finally, we present some numerical examples in Section 5.

2 Problem Formulation and Tools of Convex Optimization

In this section, we introduce the mathematical models of the generalized Sylvester problems under consideration. We also present some important tools of convex optimization used throughout the paper.

Consider the linear space \mathbb{R}^n equipped with the Euclidean norm $\|\cdot\|$. The distance function to a nonempty subset Q of \mathbb{R}^n is defined by

$$d(x; Q) := \inf\{\|x - q\| \mid q \in Q\}, \quad x \in \mathbb{R}^n. \quad (2.1)$$

Given $x \in \mathbb{R}^n$, the Euclidean projection from x to is the set

$$\Pi(x; Q) := \{q \in Q \mid d(x; Q) = \|x - q\|\}.$$

If Q is a nonempty closed convex set in \mathbb{R}^n , then $\Pi(x; Q)$ is a singleton for every $x \in \mathbb{R}^n$. Furthermore, the projection operator is non-expansive in the sense that

$$\|\Pi(x; Q) - \Pi(y; Q)\| \leq \|x - y\| \text{ for all } x, y \in \mathbb{R}^n.$$

Let Ω and Ω_i for $i = 1, \dots, m$ be nonempty closed convex subsets of \mathbb{R}^n . The mathematical modeling of the smallest intersecting ball problem with *target sets* Ω_i for $i = 1, \dots, m$ and *constraint set* Ω is

$$\text{minimize } \mathcal{D}(x) := \max\{d(x; \Omega_i) \mid i = 1, \dots, m\} \text{ subject to } x \in \Omega. \quad (2.2)$$

The solution to this problem gives the center of the smallest Euclidean ball (with center in Ω) that intersects all target sets Ω_i for $i = 1, \dots, m$.

In order to study new problems in which the intersecting Euclidean ball is replaced by balls generated by different norms, we consider a more general setting. Let F be a closed bounded convex set that contains the origin as an interior point. We hold this as our *standing assumptions* for the set F for the remainder of the paper. The *support function* associated with F is defined by

$$\sigma_F(x) := \sup\{\langle x, f \rangle \mid f \in F\}.$$

Note that if $F = \{x \in \mathbb{R}^n \mid \|x\|_X \leq 1\}$, where $\|\cdot\|_X$ is a norm in \mathbb{R}^n , then σ_F is the *dual norm* of the norm $\|\cdot\|_X$.

Let Q be a nonempty subset of \mathbb{R}^n . The generalized distance from a point $x \in \mathbb{R}^n$ to Q generated by F is given by

$$d_F(x; Q) := \inf\{\sigma_F(x - q) \mid q \in Q\}. \quad (2.3)$$

The generalized distance function (2.3) reduces to the distance function (2.1) when F is the closed unit ball of \mathbb{R}^n with respect to the Euclidean norm. The readers are referred to [14] for important properties of the generalized distance function (2.3).

Using (2.3), a more general model of problem (2.2) is given by

$$\text{minimize } \mathcal{D}_F(x) := \max\{d_F(x; \Omega_i) \mid i = 1, \dots, m\} \text{ subject to } x \in \Omega. \quad (2.4)$$

The function \mathcal{D}_F as well as its specification \mathcal{D} are nonsmooth in general. Thus, problem (2.4) and, in particular, problem (2.2) must be studied from both theoretical and numerical view points using the tools of generalized differentiation from convex analysis.

Given a function $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$, we say that φ is convex if it satisfies

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y),$$

for all $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. The function φ is said to be strictly convex if the above inequality becomes strict whenever $x \neq y$.

The class of convex functions plays an important role in many applications of mathematics, especially applications to optimization. It is well-known that for a convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, the function has an absolute minimum on a convex set Ω at \bar{x} if and only if it has a local minimum on Ω at \bar{x} . Moreover, if $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a differentiable convex function, then $\bar{x} \in \Omega$ is a minimizer for f if and only if

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \text{ for all } x \in \Omega. \quad (2.5)$$

The readers are referred to [2, 3, 10, 15] for more complete theory of convex analysis and applications to optimization from both theoretical and numerical aspects.

3 Smoothing Techniques for Generalized Sylvester Problems

In this section, we employ the approach developed in [31] to approximate the nonsmooth optimization problem (2.4) by a smooth optimization problem which is favorable for applying available smooth numerical algorithms. The difference here is that we use generalized distances to sets instead of distances to points.

Given an element $v \in \mathbb{R}^n$, the cone generated by v is given by $\text{cone}\{v\} := \{\lambda v \mid \lambda \geq 0\}$. Let us review the following definition from [14]. We recall that F is a closed bounded convex set that contains zero in its interior, as per the standing assumptions in this paper.

Definition 3.1 *The set F is normally smooth if for every $x \in \text{bd } F$ there exists $a_x \in \mathbb{R}^n$ such that $N(x; F) = \text{cone}\{a_x\}$.*

In the theorem below, we establish the necessary and sufficient condition for the smallest intersecting ball problem (2.4) to have a unique optimal solution.

Theorem 3.2 *Suppose that F is normally smooth, all of the target sets are strictly convex, and at least one of the sets $\Omega, \Omega_1, \dots, \Omega_m$ is bounded. Then the smallest intersecting ball problem (2.4) has a unique optimal solution if and only if $\bigcap_{i=1}^m (\Omega \cap \Omega_i)$ contains at most one point.*

Proof It is clear that every point in the set $\bigcap_{i=1}^m (\Omega \cap \Omega_i)$ is a solution of (2.4). Thus, if (2.4) has a unique optimal solution we must have that $\bigcap_{i=1}^m (\Omega \cap \Omega_i)$ contains at most one point, so the necessary condition has been proven.

For the sufficient condition, assume that $\bigcap_{i=1}^m (\Omega \cap \Omega_i)$ contains at most one point. The existence of an optimal solution is guaranteed by the assumption that at least one of the sets $\Omega, \Omega_1, \dots, \Omega_m$ is bounded. What remains to be shown is the uniqueness of this solution. We consider two cases.

In the first case, we assume that $\bigcap_{i=1}^m (\Omega \cap \Omega_i)$ contains exactly one point \bar{x} . Observe that $\mathcal{D}_F(\bar{x}) = 0$ and $\mathcal{D}_F(x) \geq 0$ for all $x \in \mathbb{R}^n$, so \bar{x} is a solution of (2.4). If $\hat{x} \in \Omega$ is another solution then we must have $\mathcal{D}_F(\hat{x}) = \mathcal{D}_F(\bar{x}) = 0$. Therefore, $d_F(\hat{x}; \Omega_i) = 0$ for all $i \in \{1, \dots, m\}$ and hence $\hat{x} \in \bigcap_{i=1}^m (\Omega \cap \Omega_i) = \{\bar{x}\}$. We conclude that $\hat{x} = \bar{x}$ and the problem has unique solution in this case.

For the second case we assume that $\bigcap_{i=1}^m (\Omega \cap \Omega_i) = \emptyset$. We will show that the function

$$S(x) = \max\{(d_F(x; \Omega_1))^2, \dots, (d_F(x; \Omega_m))^2\},$$

is strictly convex on Ω . This will prove the uniqueness of the solution.

Take any $x, y \in \Omega$, $x \neq y$ and $t \in (0, 1)$. Denote $x_t := tx + (1 - t)y$. Let $i \in \{1, \dots, m\}$ such that $(d_F(x_t; \Omega_i))^2 = S(x_t)$. Let $u, v \in \Omega_i$ such that $\sigma_F(x - u) = d_F(x; \Omega_i)$ and

$\sigma_F(y - v) = d_F(y; \Omega_i)$. Then we have

$$\begin{aligned}
S(x_t) &= (d_F(x_t; \Omega_i))^2 \\
&= [d_F(tx + (1-t)y; \Omega_i)]^2 \\
&\leq [td_F(x; \Omega_i) + (1-t)d_F(y; \Omega_i)]^2 \\
&= [t\sigma_F(x - u) + (1-t)\sigma_F(y - v)]^2 \\
&= t^2 (\sigma_F(x - u))^2 + 2t(1-t)\sigma_F(x - u)\sigma_F(y - v) + (1-t)^2 (\sigma_F(y - v))^2 \\
&\leq t^2 (\sigma_F(x - u))^2 + t(1-t) \left[(\sigma_F(x - u))^2 + (\sigma_F(y - v))^2 \right] + (1-t)^2 (\sigma_F(y - v))^2 \\
&= t (\sigma_F(x - u))^2 + (1-t) (\sigma_F(y - v))^2 \\
&= t (d_F(x; \Omega_i))^2 + (1-t) (d_F(y; \Omega_i))^2 \\
&\leq tS(x) + (1-t)S(y).
\end{aligned}$$

Recall that we need to prove the inequality $S(x_t) < tS(x) + (1-t)S(y)$. Suppose by contradiction that $S(x_t) = tS(x) + (1-t)S(y)$. Then all of the inequalities in the above estimate are turned to equalities and thus we have

$$d_F(x_t; \Omega_i) = td_F(x; \Omega_i) + (1-t)d_F(y; \Omega_i)$$

and

$$\sigma_F(x - u) = \sigma_F(y - v). \quad (3.6)$$

Hence,

$$d_F(x_t; \Omega_i) = \sigma_F(x - u) = \sigma_F(y - v). \quad (3.7)$$

Observe that $\sigma_F(w) = 0$ if and only if $w = 0$, (3.7) implies $x = u$ if and only if $y = v$. We claim that $x \neq u$ and $y \neq v$. Indeed, if $x = u$ and $y = v$, then $x, y \in \Omega_i$ and hence $x_t \in \Omega_i$ by the convexity of Ω_i . Thus $d_F(x_t; \Omega_i) = 0$. This contradicts the fact that $d_F(x_t; \Omega_i) = \mathcal{D}_F(x_t) > 0$ which guaranteed by the assumption $\bigcap_{i=1}^m (\Omega \cap \Omega_i) = \emptyset$.

Now, we will show that $u \neq v$. Denote $c := tu + (1-t)v \in \Omega_i$. The properties of the support function and (3.6) give

$$\begin{aligned}
d_F(x_t; \Omega_i) &\leq \sigma_F(x_t - c) \\
&= \sigma_F(t(x - u) + (1-t)(y - v)) \\
&\leq \sigma_F(t(x - u)) + \sigma_F((1-t)(y - v)) \\
&\leq t\sigma_F(x - u) + (1-t)\sigma_F(y - v) \\
&= \sigma_F(x - u).
\end{aligned}$$

By (3.7) we have $\sigma_F(t(x - u) + (1-t)(y - v)) = \sigma_F(t(x - u)) + \sigma_F((1-t)(y - v))$. Since F is normally smooth, it follows from [14, Remark 3.4] that there exists $\lambda > 0$ satisfying

$$t(x - u) = \lambda(1-t)(y - v).$$

Now, by contradiction, suppose $u = v$. Then $x - u = \beta(y - u)$ where $\beta = t^{-1}\lambda(1-t)$. Note that $\beta \neq 1$ since $x \neq y$ and $\sigma_F(y - u) > 0$ since $y - u \neq 0$. Now we have

$$\begin{aligned}
\sigma_F(x - u) &= \sup\{\langle x - u, f \rangle \mid f \in F\} = \sup\{\langle \beta(y - u), f \rangle \mid f \in F\} \\
&= \beta \sup\{\langle y - u, f \rangle \mid f \in F\} = \beta \sigma_F(y - u) \neq \sigma_F(y - u),
\end{aligned}$$

which contradicts (3.7). Thus $u \neq v$.

Since $u, v \in \Omega_i$, $u \neq v$ and Ω_i is strictly convex, $c \in \text{int } \Omega_i$. The assumption $\bigcap_{i=1}^m (\Omega \cap \Omega_i) = \emptyset$ gives $d_F(x_t; \Omega_i) = \mathcal{D}_F(x_t) > 0$. Therefore, $x_t \notin \Omega_i$ and thus $x_t \neq c$. Let $\delta > 0$ such that $\mathcal{B}(c; \delta) \subset \Omega_i$. Then $c + \gamma(x_t - c) \in \Omega_i$, with $\gamma = \frac{\delta}{2\|x_t - c\|} > 0$. We have

$$\begin{aligned} d_F(x_t; \Omega_i) &\leq \sigma_F(x_t - c - \gamma(x_t - c)) = (1 - \gamma)\sigma_F(t(x - u) + (1 - t)(y - v)) \\ &< \sigma_F(t(x - u) + (1 - t)(y - v)) = \sigma_F(x - u). \end{aligned}$$

This contradicts (3.7) and completes the proof. \square

Recall the following definition.

Definition 3.3 *A convex set F is said to be normally round if $N(x; F) \neq N(y; F)$ whenever $x, y \in \text{bd } F$, $x \neq y$.*

Proposition 3.4 *Let Θ be a nonempty closed convex subset of \mathbb{R}^n . Suppose that F is normally smooth and normally round. Then the function $g(x) := [d_F(x; \Theta)]^2$, $x \in \mathbb{R}^n$, is continuously differentiable.*

Proof It suffices to show that $\partial g(\bar{x})$ is a singleton for every $\bar{x} \in \mathbb{R}^n$. By [15], we have

$$\partial g(\bar{x}) = 2d_F(\bar{x}; \Theta)\partial d_F(\bar{x}; \Theta).$$

It follows from [14, Proposition 4.3 (iii)] that g is continuously differentiable on Θ^c , and so

$$\partial g(\bar{x}) = 2d_F(\bar{x}; \Theta)\nabla d_F(\bar{x}; \Theta) = 2d_F(\bar{x}; \Theta)\nabla\sigma_F(\bar{x} - w),$$

where $w := \Pi_F(\bar{x}; \Theta)$ and $\bar{x} \notin \Theta$.

In the case where $\bar{x} \in \Theta$, one has $d_F(\bar{x}; \Theta) = 0$, and hence

$$\partial g(\bar{x}) = 2d_F(\bar{x}; \Theta)\partial d_F(\bar{x}; \Theta) = \{0\}.$$

The proof is now complete. \square

If all of the target sets have a common point which belongs to the constraint set, then such a point is a solution of problem (2.4), so we always assume that $\bigcap_{i=1}^n (\Omega_i \cap \Omega) = \emptyset$. We also assume that at least one of the sets $\Omega, \Omega_1, \dots, \Omega_m$ is bounded which guarantees the existence of an optimal solution; see [16]. These are our *standing assumptions* for the remainder of this section.

Let us start with some useful well-known results. We include the proofs for the convenience of the reader.

Lemma 3.5 *Given positive numbers a_i for $i = 1, \dots, m$, $m > 1$, and $0 < s < t$, one has*

- (i) $(a_1^s + a_2^s + \dots + a_m^s)^{1/s} > (a_1^t + a_2^t + \dots + a_m^t)^{1/t}$.
- (ii) $(a_1^{1/s} + a_2^{1/s} + \dots + a_m^{1/s})^s < (a_1^{1/t} + a_2^{1/t} + \dots + a_m^{1/t})^t$.
- (iii) $\lim_{r \rightarrow 0^+} (a_1^{1/r} + a_2^{1/r} + \dots + a_m^{1/r})^r = \max\{a_1, \dots, a_m\}$.

Proof (i) Since $t/s > 1$, it is obvious that

$$\left(\frac{a_1^s}{\sum_{i=1}^m a_i^s}\right)^{t/s} + \cdots + \left(\frac{a_m^s}{\sum_{i=1}^m a_i^s}\right)^{t/s} < \frac{a_1^s}{\sum_{i=1}^m a_i^s} + \cdots + \frac{a_m^s}{\sum_{i=1}^m a_i^s} = 1,$$

since $\frac{a_i^s}{\sum_{i=1}^m a_i^s} \in (0, 1)$. It follows that

$$\frac{a_1^t}{(\sum_{i=1}^m a_i^s)^{t/s}} + \cdots + \frac{a_m^t}{(\sum_{i=1}^m a_i^s)^{t/s}} < 1,$$

and hence

$$a_1^t + \cdots + a_m^t < \left(\sum_{i=1}^m a_i^s\right)^{t/s}.$$

This implies **(i)** by raising both sides to the power of $1/t$.

(ii) Inequality **(ii)** follows directly from **(i)**.

(iii) Defining $a := \max\{a_1, \dots, a_m\}$ yields

$$a \leq (a_1^{1/r} + a_2^{1/r} + \dots + a_m^{1/r})^r \leq m^r a \rightarrow a \text{ as } r \rightarrow 0^+,$$

which implies **(iii)** and completes the proof. \square

For $p > 0$ and for $x \in \mathbb{R}^n$, the *log-exponential smoothing function* of $\mathcal{D}_F(x)$ is defined as

$$\mathcal{D}_F(x, p) := p \ln \sum_{i=1}^m \exp\left(\frac{G_{F,i}(x, p)}{p}\right), \quad (3.8)$$

where

$$G_{F,i}(x, p) := \sqrt{d_F(x; \Omega_i)^2 + p^2}.$$

Theorem 3.6 *The function $\mathcal{D}_F(x, p)$ defined in (3.8) has the following properties:*

(i) *If $x \in \mathbb{R}^n$ and $0 < p_1 < p_2$, then*

$$\mathcal{D}_F(x, p_1) < \mathcal{D}_F(x, p_2).$$

(ii) *For any $x \in \mathbb{R}^n$ and $p > 0$,*

$$0 \leq \mathcal{D}_F(x, p) - \mathcal{D}_F(x) \leq p(1 + \ln m).$$

(iii) *For any $p > 0$, the function $\mathcal{D}_F(\cdot, p)$ is convex. If we suppose further that F is normally smooth and the sets Ω_i for $i = 1, \dots, m$ are strictly convex and not collinear (i.e., it is impossible to draw a straight line that intersects all the sets Ω_i), then $\mathcal{D}_F(\cdot, p)$ is strictly convex.*

(iv) *For any $p > 0$, if F is normally smooth and normally round, then $\mathcal{D}_F(\cdot, p)$ is continuously differentiable.*

(v) *If at least one of the target sets Ω_i for $i = 1, \dots, m$ is bounded, then $\mathcal{D}_F(\cdot, p)$ is coercive in the sense that*

$$\lim_{\|x\| \rightarrow \infty} \mathcal{D}_F(x, p) = \infty.$$

Proof (i) Define

$$\begin{aligned} a_i(x, p) &:= \exp(G_{F,i}(x, p)), \\ a_\infty(x, p) &:= \max_{i=1, \dots, m} a_i(x, p), \text{ and} \\ G_{F,\infty}(x, p) &:= \max_{i=1, \dots, m} G_{F,i}(x, p). \end{aligned}$$

Then $a_i(x, p)$ is strictly increasing on $(0, \infty)$ as a function of p and

$$\begin{aligned} \frac{a_i(x, p)}{a_\infty(x, p)} &= \exp(G_{F,i}(x, p) - G_{F,\infty}(x, p)) \leq 1, \\ G_{F,\infty}(x, p) &\leq \mathcal{D}_F(x) + p. \end{aligned}$$

For $0 < p_1 < p_2$, it follows from Lemma 3.5(ii) that

$$\begin{aligned} \mathcal{D}_F(x, p_1) &= \ln \left[\sum_{i=1}^m (a_i(x, p_1))^{1/p_1} \right]^{p_1} < \ln \left[\sum_{i=1}^m (a_i(x, p_1))^{1/p_2} \right]^{p_2} \\ &< \ln \left[\sum_{i=1}^m (a_i(x, p_2))^{1/p_2} \right]^{p_2} = \mathcal{D}_F(x, p_2), \end{aligned}$$

which justifies **(i)**.

(ii) It follows from (3.8) that for any $i \in \{1, \dots, m\}$, we have

$$\mathcal{D}_F(x, p) \geq p \ln \exp \left(\frac{G_{F,i}(x, p)}{p} \right) = G_{F,i}(x, p) \geq d_F(x; \Omega_i).$$

This implies $\mathcal{D}_F(x, p) \geq \mathcal{D}_F(x)$ for all $x \in \mathbb{R}^n$ and $p > 0$. Moreover,

$$\begin{aligned} \mathcal{D}_F(x, p) &= \ln a_\infty(x, p) \left[\sum_{i=1}^m \left(\frac{a_i(x, p)}{a_\infty(x, p)} \right)^{1/p} \right]^p \\ &= \ln a_\infty(x, p) + p \ln \sum_{i=1}^m \left(\frac{a_i(x, p)}{a_\infty(x, p)} \right)^{1/p} \\ &\leq G_{F,\infty}(x, p) + p \ln m \leq \mathcal{D}_F(x) + p + p \ln m. \end{aligned}$$

Thus, **(ii)** has been proved.

(iii) Given $p > 0$, the function $f_p(t) := \frac{\sqrt{t^2 + p^2}}{p}$ is increasing and convex on the interval $[0, \infty)$, and $d(\cdot; \Omega_i)$ is convex, so the function $k_i(x, p) := \frac{G_{F,i}(x, p)}{p}$ is also convex with respect to x . For any $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$, by the convexity of the function

$$u = (u_1, \dots, u_m) \rightarrow \ln \sum_{i=1}^m \exp(u_i),$$

one has

$$\begin{aligned}
\mathcal{D}_F(\lambda x + (1 - \lambda)y, p) &= p \ln \sum_{i=1}^m \exp \left[k_i(\lambda x + (1 - \lambda)y, p) \right] \\
&\leq p \ln \sum_{i=1}^m \exp \left[\lambda k_i(x, p) + (1 - \lambda)k_i(y, p) \right] \\
&\leq \lambda p \ln \sum_{i=1}^m \exp \left(k_i(x, p) \right) + (1 - \lambda) p \ln \sum_{i=1}^m \exp \left(k_i(y, p) \right) \\
&= \lambda \mathcal{D}_F(x, p) + (1 - \lambda) \mathcal{D}_F(y, p).
\end{aligned} \tag{3.9}$$

Thus, $\mathcal{D}_F(\cdot, p)$ is convex. Suppose that F is normally smooth and the sets Ω_i for $i = 1, \dots, m$ are strictly convex and not collinear, but $\mathcal{D}_F(\cdot, p)$ is not strictly convex. Then there exist $x, y \in \mathbb{R}^n$ with $x \neq y$ and $0 < \lambda < 1$ such that

$$\mathcal{D}_F(\lambda x + (1 - \lambda)y, p) = \lambda \mathcal{D}_F(x, p) + (1 - \lambda) \mathcal{D}_F(y, p).$$

Thus, all the inequalities (3.9) become equalities. Since the functions \ln, \exp are strictly increasing on $(0, \infty)$, this implies

$$k_i(\lambda x + (1 - \lambda)y, p) = \lambda k_i(x, p) + (1 - \lambda)k_i(y, p), \quad \text{for all } i = 1, \dots, m. \tag{3.10}$$

Observe that $k_i(\cdot, p) = f_p(d_F(\cdot; \Omega_i))$ and the function $f_p(\cdot)$ is strictly increasing $[0, \infty)$, it follows from (3.10) that

$$d_F(\lambda x + (1 - \lambda)y, \Omega_i) = \lambda d_F(x, \Omega_i) + (1 - \lambda) d_F(y, \Omega_i) \quad \text{for all } i = 1, \dots, m.$$

The result now follows directly from the proof of [14, Proposition 4.5].

(iv) Let $\varphi_i(x) := [d_F(x; \Omega_i)]^2$. Then φ_i is continuously differentiable by Proposition 3.4. By the chain rule, for any $p > 0$, the function $\mathcal{D}_F(x, p)$ is continuously differentiable as a function of x .

(v) Without loss of generality, we assume that Ω_1 is bounded. It then follows from **(ii)** that

$$\lim_{\|x\| \rightarrow \infty} \mathcal{D}_F(x, p) \geq \lim_{\|x\| \rightarrow \infty} \mathcal{D}_F(x) \geq \lim_{\|x\| \rightarrow \infty} d_F(x; \Omega_1) = \infty.$$

Therefore, $\mathcal{D}_F(\cdot, p)$ is coercive, which justifies **(iv)**. The proof is now complete. \square

In the next corollary, we obtain an explicit formula of the gradient of the log-exponential approximation of \mathcal{D} in the case where F is the closed unit ball of \mathbb{R}^n . For $p > 0$ and for $x \in \mathbb{R}^n$, define

$$\mathcal{D}(x, p) := p \ln \sum_{i=1}^m \exp \left(\frac{G_i(x, p)}{p} \right), \tag{3.11}$$

where

$$G_i(x, p) := \sqrt{d(x; \Omega_i)^2 + p^2}.$$

Corollary 3.7 For any $p > 0$, $\mathcal{D}(\cdot, p)$ is continuously differentiable with the gradient in x computed by

$$\nabla_x \mathcal{D}(x, p) = \sum_{i=1}^m \frac{\Lambda_i(x, p)}{G_i(x, p)} (x - \tilde{x}_i),$$

where $\tilde{x}_i := \Pi(x; \Omega_i)$, and

$$\Lambda_i(x, p) := \frac{\exp(G_i(x, p)/p)}{\sum_{i=1}^m \exp(G_i(x, p)/p)}.$$

Proof It follows from Theorem 3.6 that $\mathcal{D}(\cdot, p)$ is continuously differentiable. Let $\varphi_i(x) := [d(x; \Omega_i)]^2$. Then $\nabla \varphi_i(x) = 2(x - \tilde{x}_i)$, where $\tilde{x}_i := \Pi(x; \Omega_i)$, and hence the gradient formula for $\mathcal{D}(x, p)$ follows from the chain rule. \square

Remark 3.8 (i) To avoid working with large numbers when implementing algorithms for (2.2), we often use the identity

$$\Lambda_i(x, p) := \frac{\exp(G_i(x, p)/p)}{\sum_{i=1}^m \exp[(G_i(x, p)/p)]} = \frac{\exp[(G_i(x, p) - G_\infty(x, p)]/p)}{\sum_{i=1}^m \exp[(G_i(x, p) - G_\infty(x, p)]/p)},$$

where $G_\infty(x, p) := \max_{i=1, \dots, m} G_i(x, p)$.

(ii) In general, $\mathcal{D}(\cdot, p)$ is not strictly convex. For example, in \mathbb{R}^2 , consider the sets $\Omega_1 = \{-1\} \times [-1, 1]$ and $\Omega_2 = \{1\} \times [-1, 1]$. Then $\mathcal{D}(\cdot, p)$ takes constant value on $\{0\} \times [-1, 1]$.

An important relation between problem (2.4) and problem of minimizing the function (3.11) on Ω is given in the proposition below. Note that the assumption of the proposition involves the uniqueness of an optimal solution to problem (2.4) which is guaranteed under our standing assumptions by Theorem 3.2.

Proposition 3.9 Let $\{p_k\}$ be a sequence of positive real numbers converging to 0. For each k , let $y_k \in \arg \min_{x \in \Omega} \mathcal{D}_F(x, p_k)$. Then $\{y_k\}$ is a bounded sequence and every subsequential limit of $\{y_k\}$ is an optimal solution of problem (2.4). Suppose further that problem (2.4) has a unique optimal solution. Then $\{y_k\}$ converges to that optimal solution.

Proof First, observe that $\{y_k\}$ is well defined because of the assumption that at least one of the sets $\Omega, \Omega_1, \dots, \Omega_m$ is bounded and the coercivity of $\mathcal{D}_F(\cdot, p_k)$. By Theorem 3.6 (ii), for all $x \in \Omega$, we have

$$\mathcal{D}_F(x, p_k) \leq \mathcal{D}_F(x) + p_k(1 + \ln m) \quad \text{and} \quad \mathcal{D}_F(y_k) \leq \mathcal{D}_F(y_k, p_k) \leq \mathcal{D}_F(x, p_k).$$

Thus, $\mathcal{D}_F(y_k) \leq \mathcal{D}_F(x) + p_k(1 + \ln m)$, which implies the bounded property of $\{y_k\}$ using the boundedness of Ω or the coercivity of $\mathcal{D}_F(\cdot)$ from Theorem 3.6 (v). Suppose that the subsequence $\{y_{k_l}\}$ converges to y_0 . Then $\mathcal{D}_F(y_0) \leq \mathcal{D}_F(x)$ for all $x \in \Omega$, and hence y_0 is an optimal solution of problem (2.4). If (2.4) has a unique optimal solution \bar{y} , then $y_0 = \bar{y}$ and hence $y_k \rightarrow \bar{y}$. \square

Recall that a function $\varphi : Q \rightarrow \mathbb{R}$ is called *strongly convex* with modulus $m > 0$ on a convex set Q if $\varphi(x) - \frac{m}{2}\|x\|^2$ is a convex function on Q . From the definition, it is obvious that any strongly convex function is also strictly convex. Moreover, when φ is twice differentiable, φ is strongly convex with modulus m on an open convex set Q if $\nabla^2\varphi(x) - mI$ is positive semidefinite for all $x \in Q$; see [10, Theorem 4.3.1(iii)].

Proposition 3.10 *Suppose that all the sets Ω_i for $i = 1, \dots, m$ reduce to singletons. Then for any $p > 0$, the function $\mathcal{D}(\cdot, p)$ is strongly convex on any bounded convex set, and $\nabla_x \mathcal{D}(\cdot, p)$ is globally Lipschitz continuous on \mathbb{R}^n with Lipschitz constant $\frac{2}{p}$.*

Proof Suppose that $\Omega_i = \{c_i\}$ for $i = 1, \dots, m$. Then

$$\mathcal{D}(x, p) = p \ln \sum_{i=1}^m \exp\left(\frac{g_i(x, p)}{p}\right),$$

and the gradient of $\mathcal{D}(\cdot, p)$ at x becomes

$$\nabla_x \mathcal{D}(x, p) = \sum_{i=1}^m \frac{\lambda_i(x, p)}{g_i(x, p)} (x - c_i),$$

where

$$g_i(x, p) := \sqrt{\|x - c_i\|^2 + p^2} \quad \text{and} \quad \lambda_i(x, p) := \frac{\exp(g_i(x, p)/p)}{\sum_{i=1}^m \exp(g_i(x, p)/p)}.$$

Let us denote

$$Q_{ij} := \frac{(x - c_i)(x - c_j)^T}{g_i(x, p)g_j(x, p)}.$$

Then

$$\nabla_x^2 \mathcal{D}(x, p) = \sum_{i=1}^m \left[\frac{\lambda_i(x, p)}{g_i(x, p)} (I_n - Q_{ii}) + \frac{\lambda_i(x, p)}{p} Q_{ii} - \sum_{j=1}^m \frac{\lambda_i(x, p)\lambda_j(x, p)}{p} Q_{ij} \right].$$

Given a positive constant K , for any $x \in \mathbb{R}^n$, $\|x\| < K$ and $z \in \mathbb{R}^n$, $z \neq 0$, one has

$$\begin{aligned} \frac{1}{g_i(x, p)} (\|z\|^2 - z^T Q_{ii} z) &\geq \frac{1}{g_i(x, p)} (\|z\|^2 - \|z\|^2 \|(x - c_i)/g_i(x, p)\|^2) \\ &= \frac{1}{\sqrt{\|x - c_i\|^2 + p^2}} \|z\|^2 \left[\frac{p^2}{\|x - c_i\|^2 + p^2} \right] \\ &\geq \|z\|^2 \left[\frac{p^2}{[2(\|x\|^2 + \|c_i\|^2) + p^2]^{3/2}} \right] \\ &\geq \ell \|z\|^2, \end{aligned}$$

where

$$\ell := \frac{p^2}{[2K + 2 \max_{1 \leq i \leq m} \|c_i\|^2 + p^2]^{3/2}}.$$

For m real numbers a_1, \dots, a_m , since $\lambda_i(x, p) \geq 0$ for all $i = 1, \dots, m$, and $\sum_{i=1}^m \lambda_i(x, p) = 1$, by Cauchy-Schwartz inequality, we have

$$\left(\sum_{i=1}^m \lambda_i(x, p) a_i \right)^2 = \left(\sum_{i=1}^m \sqrt{\lambda_i(x, p)} \sqrt{\lambda_i(x, p)} a_i \right)^2 \leq \sum_{i=1}^m \lambda_i(x, p) a_i^2.$$

This implies

$$\begin{aligned} z^T \nabla_x^2 \mathcal{D}(x, p) z &= \sum_{i=1}^m \left[\frac{\lambda_i(x, p)}{g_i(x, p)} (\|z\|^2 - z^T Q_{ii} z) \right] \\ &\quad + \sum_{i=1}^m \left[\frac{\lambda_i(x, p)}{p} z^T Q_{ii} z - \sum_{j=1}^m \frac{\lambda_i(x, p) \lambda_j(x, p)}{p} z^T Q_{ij} z \right]. \\ &\geq \ell \|z\|^2 + \frac{1}{p} \left[\sum_{i=1}^m \lambda_i(x, p) a_i^2 - \left(\sum_{i=1}^m \lambda_i(x, p) a_i \right)^2 \right] \\ &\geq \ell \|z\|^2, \end{aligned}$$

where $a_i := z^T(x - c_i)/g_i(x, p)$. This shows that $\mathcal{D}(x, p)$ is strongly convex on $B(0; K)$.

The fact that for any $p > 0$, the gradient of $\mathcal{D}(x, p)$ with respect to x is Lipschitz continuous with constant $L = \frac{2}{p}$ was proved in [29, Proposition 2]. \square

4 The Minimization Majorization Algorithm for Generalized Sylvester Problems

In this section, we apply the minimization majorization well known in computational statistics along with the log-exponential smoothing technique developed in the previous section to develop an algorithm for solving the smallest intersecting ball problem. We also provide some examples showing that minimizing functions that involve distances to convex sets not only allows to study a generalized version of the smallest enclosing circle problem, but also opens up the possibility of applications to other problems of constrained optimization.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. Consider the optimization problem

$$\text{minimize } f(x) \text{ subject to } x \in \Omega. \tag{4.12}$$

A function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a *surrogate* of f at $\bar{z} \in \Omega$ if

$$\begin{aligned} f(x) &\leq g(x) \text{ for all } x \in \Omega, \\ f(\bar{z}) &= g(\bar{z}). \end{aligned}$$

The set of all surrogates of f at \bar{z} is denoted by $S(f, \bar{z})$.

The minimization majorization algorithm for solving (4.12) is given as follows; see [13].

Algorithm 1.

```

INPUT:  $x_0 \in \Omega, N$ 
for  $k = 1, \dots, N$  do
    Find a surrogate  $g_k \in S(f, x_{k-1})$ 
    Find  $x_k \in \operatorname{argmin}_{x \in \Omega} g_k(x)$ 
end for
OUTPUT:  $x_N$ 

```

Clearly, the choice of surrogate $g_k \in S(f, x_{k-1})$ plays a crucial role in the minimization majorization algorithm. In what follows, we consider a particular choice of surrogates for the minimization majorization algorithm; see, e.g., [5, 11, 12]. An objective function $f : \Omega \rightarrow \mathbb{R}$ is said to be *majorized* by $\mathcal{M} : \Omega \times \Omega \rightarrow \mathbb{R}$ if

$$f(x) \leq \mathcal{M}(x, y) \text{ and } \mathcal{M}(y, y) = f(y) \text{ for all } x, y \in \Omega.$$

Given $x_{k-1} \in \Omega$, we can define $g_k(x) := \mathcal{M}(x, x_{k-1})$, so that $g_k \in S(f, x_{k-1})$. Then the update

$$x_k \in \arg \min_{x \in \Omega} \mathcal{M}(x, x_{k-1})$$

defines an minimization majorization algorithm. As mentioned above, finding an appropriate majorization is an important piece of this algorithm. It has been shown in [5] that the minimization majorization algorithm using distance majorization provides an effective tool for solving many important classes of optimization problems. The key step is to use the following:

$$d(x; Q) \leq \|x - \Pi(y; Q)\| \text{ and } d(y; Q) = \|y - \Pi(y; Q)\|.$$

In the examples below, we revisit some algorithms based on distance majorization and provide the convergence analysis for these algorithms.

Example 4.1 Let Ω_i for $i = 1, \dots, m$ be nonempty closed convex subsets of \mathbb{R}^n such that $\bigcap_{i=1}^m \Omega_i \neq \emptyset$. The problem of finding a point $x^* \in \bigcap_{i=1}^m \Omega_i$ is called the *feasible point problem* for these sets. Consider the problem

$$\text{minimize } f(x) := \sum_{i=1}^m [d(x; \Omega_i)]^2, x \in \mathbb{R}^n. \quad (4.13)$$

With the assumption that $\bigcap_{i=1}^m \Omega_i \neq \emptyset$, $x^* \in \mathbb{R}^n$ is an optimal solution of (4.13) if and only if $x^* \in \bigcap_{i=1}^m \Omega_i$. Thus, we only need to consider (4.13).

Let us apply the minimization majorization algorithm for (4.13). First, we need to find surrogates for the objective function $f(x) = \sum_{i=1}^m [d(x; \Omega_i)]^2$. Let

$$g_k(x) := \sum_{i=1}^m \|x - \Pi(x_{k-1}; \Omega_i)\|^2.$$

Then $g_k \in \mathcal{S}(f, x_{k-1})$ for all $k \in \mathbb{N}$, so the minimization majorization algorithm is given by

$$x_k \in \operatorname{argmin}_{x \in \mathbb{R}^n} g_k(x) = \frac{1}{m} \sum_{i=1}^m \Pi(x_{k-1}; \Omega_i).$$

Let

$$h_k(x) := g_k(x) - f(x) = \sum_{i=1}^m [\|x - \Pi(x_{k-1}; \Omega_i)\|^2 - [d(x; \Omega_i)]^2].$$

We can show that $h_k(x)$ is differentiable on \mathbb{R}^n and $\nabla h_k(x)$ is Lipschitz with constant $L = 2m$. Moreover,

$$h_k(x_{k-1}) = \nabla h_k(x_{k-1}) = 0.$$

The function $g_k(x) - m\|x\|^2$ is convex, so g_k is strongly convex with modulus $\rho = 2m$. Using the same notation as in [13], one has $g_k \in \mathcal{S}_{L,\rho}(f, x_{k-1})$ with $\rho = L = 2m$. By [13, Proposition 2.8],

$$f(x_k) - V_* \leq \frac{m}{k} \|x_0 - x^*\|^2 \text{ for all } k \in \mathbb{N}.$$

Example 4.2 Given a *data set* $S := \{(a_i, y_i)\}_{i=1}^m$, where $a_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$, consider the *support vector machine* problem

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|x\|^2 \\ & \text{subject to } y_i \langle a_i, x \rangle \geq 1 \text{ for all } i = 1, \dots, m. \end{aligned}$$

Let $\Omega_i := \{x \in \mathbb{R}^p \mid y_i \langle a_i, x \rangle \geq 1\}$. Using the quadratic penalty method (see [5]), the support vector machine problem can be solved by the following unconstrained optimization problem:

$$\text{minimize } f(x) := \frac{1}{2} \|x\|^2 + \frac{C}{2} \sum_{i=1}^m [d(x; \Omega_i)]^2, x \in \mathbb{R}^p, C > 0. \quad (4.14)$$

Using the minimization majorization algorithm with the surrogates

$$g_k(x) = \frac{1}{2} \|x\|^2 + \frac{C}{2} \sum_{i=1}^m \|x - \Pi(x_{k-1}; \Omega_i)\|^2$$

for (4.14) yields

$$x_k = \frac{C}{1 + mC} \sum_{i=1}^m \Pi(x_{k-1}, \Omega_i).$$

Let

$$h_k(x) := g_k(x) - f(x) = \frac{C}{2} \sum_{i=1}^m [\|x - \Pi(x_{k-1}; \Omega_i)\|^2 - [d(x; \Omega_i)]^2].$$

We can show that $\nabla h_k(x)$ is Lipschitz with constant $L = mC$, and $h_k(x_{k-1}) = \nabla h_k(x_{k-1}) = 0$. Moreover, g_k is strongly convex with parameter $\rho = 1 + mC$. By [13, Proposition 2.8], the minimization majorization method applied for (4.14) gives

$$f(x_k) - V_* \leq \frac{mC}{2} \left(\frac{mC}{mC + 2} \right)^{k-1} \|x_0 - x^*\|^2 \text{ for all } k \in \mathbb{N},$$

where x_* is the optimal solution of (4.14) and V_* is the optimal value.

In what follows, we apply the minimization majorization algorithm in combination with the log-exponential smoothing technique to solve the smallest intersecting ball problem (2.2). In the first step, we approximate the cost function \mathcal{D} in (2.2) by the log-exponential smoothing function (3.11). Then the new function is majorized in order to apply the minimization majorization algorithm. For $x, y \in \mathbb{R}^n$ and $p > 0$, define

$$\mathcal{G}(x, y, p) := p \ln \sum_{i=1}^m \exp \left(\frac{\sqrt{\|x - \Pi(y; \Omega_i)\|^2 + p^2}}{p} \right).$$

Then $\mathcal{G}(x, y, p)$ serves as a majorization of the log-exponential smoothing function (3.11). From Proposition 3.10, for $p > 0$ and $y \in \mathbb{R}^n$, the function $\mathcal{G}(x, y, p)$ with variable x is strongly convex on any bounded set and continuously differentiable with Lipschitz gradient on \mathbb{R}^n .

Our algorithm is explained as follows. Choose a small number \bar{p} . In order to solve the smallest intersecting ball problem (2.2), we minimize its log-exponential smoothing approximation (3.11):

$$\text{minimize } \mathcal{D}(x, \bar{p}) \text{ subject to } x \in \Omega. \quad (4.15)$$

Pick an initial point $x_0 \in \Omega$ and apply the minimization majorization algorithm with

$$x_k := \arg \min_{x \in \Omega} \mathcal{G}(x, x_{k-1}, \bar{p}). \quad (4.16)$$

The algorithm is summarized by the following.

Algorithm 2.

```

INPUT:  $\Omega, \bar{p} > 0, x_0 \in \Omega, m$  target sets  $\Omega_i, i = 1, \dots, m, N$ 
for  $k = 1, \dots, N$  do
    use a fast gradient algorithm to solve approximately
         $x_k := \arg \min_{x \in \Omega} \mathcal{G}(x, x_{k-1}, \bar{p})$ 
end for
OUTPUT:  $x_N$ 

```

Proposition 4.3 *Given $\bar{p} > 0$ and $x_0 \in \Omega$, the sequence $\{x_k\}$ of exact solutions $x_k := \arg \min_{x \in \Omega} \mathcal{G}(x, x_{k-1}, \bar{p})$ generated by Algorithm 2 has a convergent subsequence.*

Proof Denoting $\alpha := \mathcal{D}(x_0, \bar{p})$ and using Theorem 3.6(v) imply that the level set

$$\mathcal{L}_{\leq \alpha} := \{x \in \Omega \mid \mathcal{D}(x, \bar{p}) \leq \alpha\}$$

is bounded. For any $k \geq 1$, because $\mathcal{G}(\cdot, x_{k-1}, \bar{p})$ is a surrogate of $\mathcal{D}(\cdot, \bar{p})$ at x_{k-1} , one has

$$\mathcal{D}(x_k, \bar{p}) \leq \mathcal{G}(x_k, x_{k-1}, \bar{p}) \leq \mathcal{G}(x_{k-1}, x_{k-1}, \bar{p}) = \mathcal{D}(x_{k-1}, \bar{p}).$$

It follows that

$$\mathcal{D}(x_k, \bar{p}) \leq \mathcal{D}(x_{k-1}, \bar{p}) \leq \dots \leq \mathcal{D}(x_1, \bar{p}) \leq \mathcal{D}(x_0, \bar{p}).$$

This implies $\{x_k\} \subset \mathcal{L}_{\leq \alpha}$ which is a bounded set, so $\{x_k\}$ has a convergent subsequence. \square

The convergence of the minimization majorization algorithm depends on the algorithm map

$$\psi(x) := \arg \min_{y \in \Omega} \mathcal{G}(y, x, \bar{p}) = \arg \min_{y \in \Omega} \left\{ \bar{p} \ln \sum_{i=1}^m \exp \left(\frac{\sqrt{\|y - \Pi(x; \Omega_i)\|^2 + \bar{p}^2}}{\bar{p}} \right) \right\}. \quad (4.17)$$

In the theorem below, we show that the conditions in [5, Proposition 1] are satisfied.

Theorem 4.4 *Given $\bar{p} > 0$, the function $\mathcal{D}(\cdot, \bar{p})$ and the algorithm map $\psi : \Omega \rightarrow \Omega$ defined by (4.17) satisfy the following conditions:*

(i) *For any $x_0 \in \Omega$, the level set*

$$\mathcal{L}(x_0) := \{x \in \Omega \mid \mathcal{D}(x, \bar{p}) \leq \mathcal{D}(x_0, \bar{p})\}$$

is compact.

(ii) *ψ is continuous on Ω .*

(iii) *$\mathcal{D}(\psi(x), \bar{p}) < \mathcal{D}(x, \bar{p})$ whenever $x \neq \psi(x)$.*

(iv) *Any fixed point \bar{x} of ψ is a minimizer of $\mathcal{D}(\cdot, \bar{p})$ on Ω .*

Proof Observe that the function $\mathcal{D}(\cdot, \bar{p})$ is continuous on Ω . Then the level set $\mathcal{L}(x_0)$ is compact for any initial point x_0 since $\mathcal{D}(\cdot, \bar{p})$ is coercive by Theorem 3.6(v), and hence (i) is satisfied. From the strict convexity on Ω of $\mathcal{G}(\cdot, x, \bar{p})$ guaranteed by Proposition 3.10, we can show that the algorithm map $\psi : \Omega \rightarrow \Omega$ is a single-valued mapping. Let us prove that ψ is continuous. Take an arbitrary sequence $\{x_k\} \subset \Omega$, $x_k \rightarrow \bar{x} \in \Omega$ as $k \rightarrow \infty$. It suffices to show that the sequence $y_k := \psi(x_k)$ tends to $\psi(\bar{x})$. It follows from the continuity of $\mathcal{D}(\cdot, \bar{p})$ that $\mathcal{D}(x_k, \bar{p}) \rightarrow \mathcal{D}(\bar{x}, \bar{p})$, and hence we can assume

$$\mathcal{D}(x_k, \bar{p}) \leq \mathcal{D}(\bar{x}, \bar{p}) + \delta,$$

for all $k \in \mathbb{N}$, where δ is a positive constant. One has the estimates

$$\mathcal{D}(\psi(x_k), \bar{p}) \leq \mathcal{G}(\psi(x_k), x_k, \bar{p}) \leq \mathcal{G}(x_k, x_k, \bar{p}) = \mathcal{D}(x_k, \bar{p}) \leq \mathcal{D}(\bar{x}, \bar{p}) + \delta,$$

which imply that $\{y_k\}$ is bounded by the coerciveness of $\mathcal{D}(\cdot, \bar{p})$. Consider any convergent subsequence $\{y_{k_\ell}\}$ with the limit z . Since y_{k_ℓ} is a solution of the smooth optimization problem $\min_{y \in \Omega} \mathcal{G}(y, x_{k_\ell}, \bar{p})$, by the necessary and sufficient optimality condition (2.5) for the given smooth convex constrained optimization problem, we have

$$\langle \nabla \mathcal{G}(y_{k_\ell}, x_{k_\ell}, \bar{p}), x - y_{k_\ell} \rangle \geq 0 \text{ for all } x \in \Omega.$$

This is equivalent to

$$\left\langle \sum_{i=1}^m \frac{\lambda_i(y_{k_\ell}, \bar{p})}{g_i(y_{k_\ell}, \bar{p})} (y_{k_\ell} - \Pi(x_{k_\ell}; \Omega_i)), x - y_{k_\ell} \right\rangle \geq 0 \text{ for all } x \in \Omega,$$

where

$$g_i(y_{k_\ell}, \bar{p}) = \sqrt{\|y_{k_\ell} - \Pi(x_{k_\ell}; \Omega_i)\|^2 + \bar{p}^2} \text{ and } \lambda_i(y_{k_\ell}, \bar{p}) = \frac{\exp(g_i(y_{k_\ell}, \bar{p})/\bar{p})}{\sum_{i=1}^m \exp(g_i(y_{k_\ell}, \bar{p})/\bar{p})}.$$

Since the Euclidean projection mapping to a nonempty closed convex set is continuous, by passing to a limit, we have

$$\langle \nabla \mathcal{G}(z, \bar{x}, \bar{p}), x - z \rangle \geq 0 \text{ for all } x \in \Omega.$$

Thus, applying (2.5) again implies that z is also an optimal solution of the problem $\min_{y \in \Omega} \mathcal{G}(y, \bar{x}, \bar{p})$. By the uniqueness of solution and $\psi(\bar{x}) = \arg \min_{y \in \Omega} \mathcal{G}(y, \bar{x}, \bar{p})$, one has that $z = \psi(\bar{x})$ and $\{y_{k_\ell}\}$ converges to $\psi(\bar{x})$. Since this conclusion holds for all convergent subsequences of the bounded sequence $\{y_k\}$, the sequence $\{y_k\}$ itself converges to $\psi(\bar{x})$, which shows that **(ii)** is satisfied. Let us verify that $\mathcal{D}(\psi(x), \bar{p}) < \mathcal{D}(x, \bar{p})$ whenever $\psi(x) \neq x$. Observe that $\psi(x) = x$ if and only if $\mathcal{G}(x, x, \bar{p}) = \min_{y \in \Omega} \mathcal{G}(y, x, \bar{p})$. Since $\mathcal{G}(y, x, \bar{p})$ has a unique minimizer, we have the strict inequality $\mathcal{G}(\psi(x), x, \bar{p}) < \mathcal{G}(x, x, \bar{p})$ whenever x is not a fix point of ψ . Combining with $\mathcal{D}(\psi(x), \bar{p}) \leq \mathcal{G}(\psi(x), x, \bar{p})$ and $\mathcal{D}(x, \bar{p}) = \mathcal{G}(x, x, \bar{p})$, we arrive at the conclusion **(iii)**.

Finally, we show that, any fixed point \bar{x} of algorithm map $\psi(x)$ is a minimizer of $\mathcal{D}(x, \bar{p})$ on Ω . Fix any $\bar{x} \in \Omega$ such that $\psi(\bar{x}) = \bar{x}$. Then $\mathcal{G}(\bar{x}, \bar{x}, \bar{p}) = \min_{y \in \Omega} \mathcal{G}(y, \bar{x}, \bar{p})$, which is equivalent to

$$\langle \nabla \mathcal{G}(\bar{x}, \bar{x}, \bar{p}), x - \bar{x} \rangle \geq 0 \text{ for all } x \in \Omega.$$

This means

$$\left\langle \sum_{i=1}^m \frac{\lambda_i(\bar{x}, \bar{p})}{g_i(\bar{x}, \bar{p})} (\bar{x} - \Pi(\bar{x}; \Omega_i)), x - \bar{x} \right\rangle \geq 0 \text{ for all } x \in \Omega,$$

where

$$g_i(\bar{x}, \bar{p}) = \sqrt{\|\bar{x} - \Pi(\bar{x}, \Omega_i)\|^2 + \bar{p}^2} = \sqrt{d(\bar{x}; \Omega_i)^2 + \bar{p}^2} = G_i(\bar{x}, \bar{p})$$

and

$$\lambda_i(\bar{x}, \bar{p}) = \frac{\exp(g_i(\bar{x}, \bar{p})/\bar{p})}{\sum_{i=1}^m \exp(g_i(\bar{x}, \bar{p})/\bar{p})} = \Lambda_i(\bar{x}, \bar{p}).$$

This inequality, however, is equivalent to the inequality $\langle \nabla \mathcal{D}(\bar{x}, \bar{p}), x - \bar{x} \rangle \geq 0$ for all $x \in \Omega$, which in turn holds if and only if \bar{x} is a minimizer of $\mathcal{D}(x, \bar{p})$ on Ω . \square

Corollary 4.5 *Given $\bar{p} > 0$ and $x_0 \in \Omega$, the sequence $\{x_k\}$ of exact solution $x_k := \arg \min_{x \in \Omega} \mathcal{G}(x, x_{k-1}, \bar{p})$ generated by Algorithm 2 has a subsequence that converges to an optimal solution of (4.15). If we suppose further that problem (4.15) has a unique optimal solution, then $\{x_k\}$ converges to this optimal solution.*

Proof It follows from Proposition 4.3 that $\{x_k\}$ has a subsequence $\{x_{k_\ell}\}$ that converges to \bar{x} . Applying [5, Proposition 1] implies that $\|x_{k_\ell+1} - x_{k_\ell}\| \rightarrow 0$ as $k \rightarrow \infty$. From the continuity of the algorithm map ψ and the equation $x_{k_\ell+1} = \psi(x_{k_\ell})$, one has that $\psi(\bar{x}) = \bar{x}$. By Theorem 4.4(iv), the element \bar{x} is an optimal solution of (4.15). The last conclusion is obvious. \square

In what follows, we apply *Nesterov's accelerated gradient method* introduced in [18, 20] to solve (4.16) approximately. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a smooth convex function with Lipschitz gradient. That is, there exists $\ell \geq 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq \ell \|x - y\| \text{ for all } x, y \in \mathbb{R}^n.$$

Let Ω be a nonempty closed convex set. In his seminal papers [18, 20], Nesterov considered the optimization problem

$$\text{minimize } f(x) \text{ subject to } x \in \Omega.$$

For $x \in \mathbb{R}^n$, define

$$T_{\Omega}(x) := \arg \min \{ \langle \nabla f(x), y - x \rangle + \frac{\ell}{2} \|x - y\|^2 \mid y \in \Omega \}.$$

Let $d : \mathbb{R}^n \rightarrow \mathbb{R}$ be a strongly convex function with parameter $\sigma > 0$. Let $x_0 \in \mathbb{R}^n$ such that

$$x_0 = \arg \min \{ d(x) \mid x \in \Omega \}.$$

Further, assume that $d(x_0) = 0$.

For simplicity, we choose $d(x) = \frac{1}{2} \|x - x_0\|^2$, where $x_0 \in \Omega$, so $\sigma = 1$. It is not hard to see that

$$y_k = T_{\Omega}(x_k) = \Pi(x_k - \frac{\nabla f(x_k)}{\ell}; \Omega).$$

Moreover,

$$z_k = \Pi(x_0 - \frac{1}{\ell} \sum_{i=0}^k \frac{i+1}{2} \nabla f(x_i); \Omega).$$

Nesterov's accelerated gradient algorithm is outlined as follows.

Algorithm 3.

```

INPUT:  $f, \ell, x_0 \in \Omega$ 
set  $k = 0$ 
repeat
  find  $y_k := T_{\Omega}(x_k)$ 
  find  $z_k := \arg \min \{ \frac{\ell}{\sigma} d(x) + \sum_{i=0}^k \frac{i+1}{2} [f(x_i) + \langle \nabla f(x_i), x - x_i \rangle] \mid x \in \Omega \}$ 
  set  $x_k := \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$ 
  set  $k := k + 1$ 
until a stopping criterion is satisfied.
OUTPUT:  $y_k$ .

```

It has been experimentally observed that the algorithm is more effective if, instead of choosing a small value p ahead of time, we change its value using an initial value p_0 and define $p_s := \sigma p_{s-1}$, where $\sigma \in (0, 1)$.

Algorithm 4.

```

INPUT:  $\Omega, \epsilon > 0, p_0 > 0, x_0 \in \Omega, m$  target sets  $\Omega_i, i = 1, \dots, m, N$ 
set  $p = p_0, y = x_0$ 
for  $k = 1, \dots, N$  do
  use Nesterov's accelerated gradient method to solve approximately
   $y := \arg \min_{x \in \Omega} \mathcal{G}(x, y, p)$ 
  set  $p := \sigma p$ 
end for
OUTPUT:  $y$ 

```

Remark 4.6 (i) When implementing this algorithm, we usually desire to maintain $p_k > \epsilon$, where $\epsilon < p_0$. So the factor σ can be calculated based on the desired number of iterations N to be run, i.e., $\sigma = (\epsilon/p_0)^{1/N}$.

(ii) In Nesterov’s accelerated gradient method, at iteration k , we often use the stopping criterion

$$\|\nabla_x \mathcal{G}(x, y, p)\| < \gamma_k,$$

where γ_0 is chosen and $\gamma_k = \tilde{\sigma}\gamma_{k-1}$ for some $\tilde{\sigma} \in (0, 1)$. The factor $\tilde{\sigma}$ can be calculated based on the number of iterations N and the lower bound $\tilde{\epsilon} > 0$ for γ_k as above.

5 Numerical Implementation

We implement Algorithm 4 to solve the generalized Sylvester problem in a number of examples. In each of the following examples, we implement Algorithm 4 with the following parameters described in this algorithm and Remark 4.6: $\epsilon = 10^{-6}$, $\tilde{\epsilon} = 10^{-5}$, $p_0 = 5$, $\gamma_0 = .5$, and $N = 10$. Observations suggest that when the number of dimensions is large, speed is improved by starting with the relatively high γ_0 and p_0 and decreasing each (thereby reducing error) with each iteration. Choosing $\sigma, \tilde{\sigma}$ as described in Remark 4.6 ensures that the final iterations are of desired accuracy. The approximate radii of the smallest intersecting ball that corresponds to the approximate optimal solution x_k is $r_k := \mathcal{D}(x_k)$.

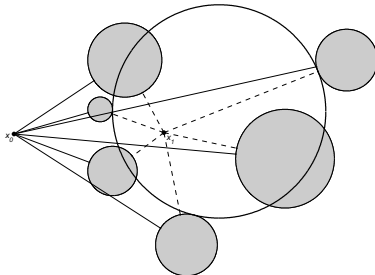


Figure 1: An illustration of minimization majorization algorithm.

Example 5.1 Let us first apply Algorithm 4 to an unconstrained generalized Sylvester problem (2.2) in \mathbb{R}^2 in which Ω_i for $i = 1, \dots, 6$ are disks with centers at $(-6, 9)$, $(12, 9)$, $(-1, -6)$, $(-8, 5)$, $(-7, 0)$, $(7, 1)$ with radii 3, 2.5, 2.5, 1, 2, 4, respectively. This setup is depicted in Figure 1. A simple MATLAB program yields an approximate smallest intersecting ball with center $x_* \approx (1.65, 4.83)$ with the approximate radius $r_* \approx 8.65$. Figure 1 shows a significant move toward the optimal solution of the problem in one step of the minimization majorization algorithm.

Example 5.2 We consider the smallest intersecting ball problem in which the target sets are square boxes in \mathbb{R}^n . In \mathbb{R}^n , a square box $S(\omega, r)$ with center $\omega = (\omega_1, \dots, \omega_n)$ and radius

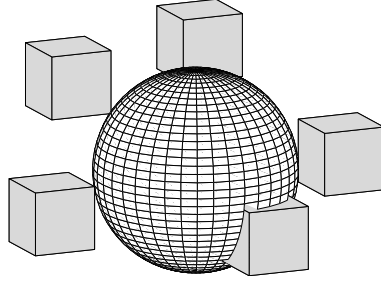


Figure 2: A smallest intersecting ball problem for cubes in \mathbb{R}^3 .

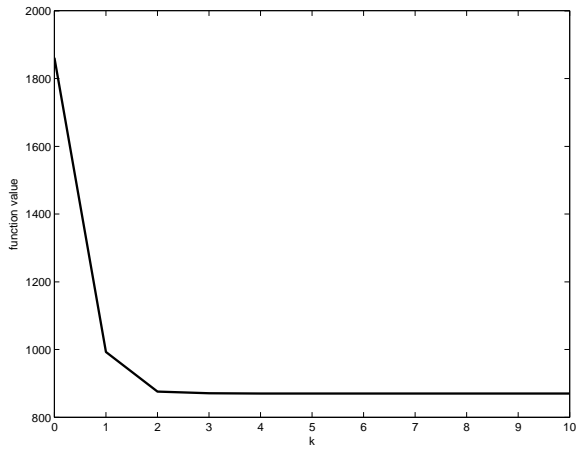
r is the set

$$S(\omega, r) := \{x = (x_1, \dots, x_n) \mid \max\{|x_1 - \omega_1|, \dots, |x_n - \omega_n|\} \leq r\}.$$

Note that the Euclidean projection from x to $S(\omega, r)$ can be expressed componentwise as follows

$$[\Pi(x; S)]_i = \begin{cases} \omega_i - r_i & \text{if } x_i + r \leq \omega_i, \\ x_i & \text{if } \omega_i - r \leq x_i \leq \omega_i + r, \\ \omega_i + r & \text{if } \omega_i + r \leq x_i. \end{cases}$$

Consider the case where $n = 3$. The target sets are 5 square boxes with centers $(-5, 0, 0)$, $(1, 4, 4)$, $(0, 5, 0)$, $(-4, -3, 2)$ and $(0, 0, 5)$ and radii $r_i = 1$ for $i = 1, \dots, 5$. Our results show that both Algorithm 4 and the subgradient method give an approximate SIB radius $r_* \approx 3.18$; see Figure 2.



MATLAB RESULT	
k	r_k
0	1861.36441
1	992.32230
2	875.27618
3	870.47714
4	869.94621
5	869.82005
6	869.79982
7	869.79676
8	869.79628
9	869.79621
10	869.79619

Figure 3: A smallest intersecting ball problem for cubes in high dimensions.

Example 5.3 Now we illustrate the performance of Algorithm 4 in high dimensions with the same setting in Example 5.2. We choose a modification of the pseudo random sequence from [31] with $a_0 = 7$ and for $i = 1, 2, \dots$

$$a_{i+1} = \text{mod}(445a_i + 1, 4096), \quad b_i = \frac{a_i}{40.96}.$$

The radii r_i and centers c_i of the square boxes are successively set to b_1, b_2, \dots in the following order

$$10r_1, c_1(1), \dots, c_1(n); 10r_2, c_2(1), \dots, c_2(n); \dots; 10r_m, c_m(1), \dots, c_m(n).$$

Consider $m = 100, n = 1000$. Figure 3 shows the approximate values of the radii r_k for $k = 0, \dots, 10$.

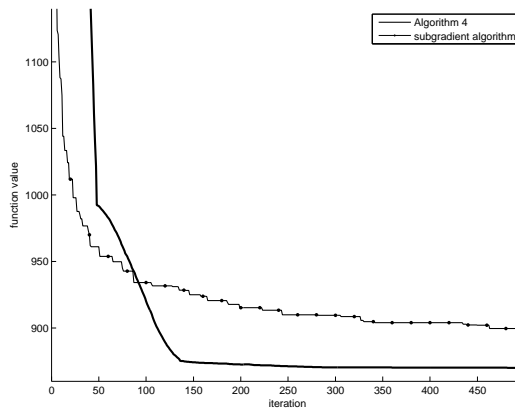


Figure 4: Comparison between minimization majorization algorithm and subgradient algorithm.

We also implement algorithm 4 in comparison with the subgradient algorithm. From our numerical results, we see that Algorithm 4 performs much better than the subgradient algorithm in both accuracy and speed. In the case where the number of target sets is large or the dimension m is high, the subgradient algorithm seems to be stagnated but Algorithm 4 still performs well. Figure 4 shows that comparison between Algorithm 4 and the subgradient algorithm. Note that in Algorithm 4, we count every iteration of Nesterov’s accelerated gradient method in the total iteration count along the horizontal axis. Thus the “sharp corner” that can be seen at 50 iterations represents the transition from x_0 to x_1 , and a subsequent recalculation of p by $p = \sigma p$.

6 Concluding Remarks

Based on the log-exponential smoothing technique and the minimization majorization algorithm, we have developed an effective numerical algorithm for solving the smallest intersecting ball problem. The problem under consideration not only generalizes the Sylvester

smallest enclosing circle problem, but also opens up the possibility of applications to other problems of constrained optimization, especially those that appear frequently in machine learning. Our numerical examples show that the algorithm works well for the problem in high dimensions. Although a number of key convergence results are contained in this paper, our future work will further develop an understanding of the convergence rate of this algorithm

Acknowledgement. The author would like to thank Prof. Jie Sun for giving comments that help improve the presentation of the paper.

References

- [1] J. Alonso, H. Martini, and M. Spirova, Minimal enclosing discs, circumcircles, and circumcenters in normed planes, *Comput. Geom.* **45** (2012), 258–274.
- [2] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Boston, 2003.
- [3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, England, 2004.
- [4] D. Cheng, X. Hu, C. Martin, On the smallest enclosing balls, *Commun. Inf. Syst.* **6** (2006), 137–160.
- [5] E. Chi, H. Zhou, K. Lange, Distance Majorization and Its Applications, *Math Program.*, in press.
- [6] L. Drager, J. Lee, C. Martin, On the geometry of the smallest circle enclosing a finite set of points, *J. Franklin Inst.* **344** (2007), 929–940.
- [7] K. Fischer and B. Gartner, The smallest enclosing ball of balls: Combinatorial structure and algorithms, *Comput. Geom.* **14** (2004), 341–378.
- [8] S. K. Jacobsen, An algorithm for the minimax Weber problem, *European J. Oper. Res.* **6** (1981), 144–148.
- [9] D. W. Hearn and J. Vijay, Efficient algorithms for the (weighted) minimum circle problem, *Oper. Res.* **30** (1981), 777–795.
- [10] J. B. Hiriart-Urruty and C. Lemaréchal, *Fundamental of Convex Analysis*, Springer-Verlag, 2001.
- [11] D.R. Hunter, K. Lange, Tutorial on MM algorithms. *Amer Statistician* **58**(2004), 30–37.
- [12] K. Lange, D. R. Hunter, I. Yang, Optimization transfer using surrogate objective functions (with discussion). *J Comput Graphical Statistis* **9** (2000), 1–59.

- [13] J. Mairal. Incremental majorization-minimization optimization with application to large-scale machine learning. arXiv preprint arXiv:1402.4419 (2014).
- [14] N. M. Nam, N. T. An, R. B. Rector and J. Sun, Nonsmooth algorithms and Nesterov’s smoothing technique for generalized Fermat-Torricelli problems. *SIAM J. Optim.* **24** (2014), no. 4, 1815–1839.
- [15] B. S. Mordukhovich and N. M. Nam, *An Easy Path to Convex Analysis and Applications*, Morgan & Claypool Publishers, 2014.
- [16] N. M. Nam and N. Hoang, A generalized Sylvester problem and a generalized Fermat-Torricelli problem, to appear in *Journal of Convex Analysis*.
- [17] N. M. Nam, N. T. An and J. Salinas, Applications of convex analysis to the smallest intersecting ball problem, *J. Convex Anal.* **19** (2012), 497–518.
- [18] Yu. Nesterov, Smooth minimization of non-smooth functions, *Math. Program.* **103** (2005), 127–152.
- [19] Yu. Nesterov, *Introductory lectures on convex optimization. A basic course. Applied Optimization*, 87. Kluwer Academic Publishers, Boston, MA, 2004.
- [20] Yu. Nesterov, A method for unconstrained convex minimization problem with the rate of convergence $O(\frac{1}{k^2})$. *Doklady AN SSSR* (translated as *Soviet Math. Docl.*) **269**(1983), 543–547.
- [21] F. Nielsen and R. Nock, Approximating smallest enclosing balls with applications to machine learning, *Internat. J. Comput. Geom. Appl.* **19** (2009), 389–414.
- [22] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [23] A. Saha, S. Vishwanathan, and X. Zhang, Efficient approximation algorithms for minimum enclosing convex shapes, *proceedings of SODA*, 2011.
- [24] J. J. Sylvester, A question in the geometry of situation. *Quarterly Journal of Pure and Applied Mathematics* **1:79**(1857).
- [25] E. Welzl, Smallest enclosing disks (balls ellipsoids). H. Maurer, editor, *Lecture Notes in Comput. Sci.* **555** (1991), 359–370.
- [26] S. Xu, R. M. Freund, and J. Sun, Solution methodologies for the smallest enclosing circle problem. A tribute to Elijah (Lucien) Polak. *Comput. Optim. Appl.* **25**(2003), no. 1-3, 283–292.
- [27] E. A. Yildirim, On the minimum volume covering ellipsoid of ellipsoids, *SIAM J. Optim.* **17** (2006), 621–641.
- [28] E. A. Yildirim, Two algorithms for the minimum enclosing ball problem, *SIAM J. Optim.* **19** (2008), 1368–1391.

- [29] X. Zhai, Two problems in convex conic optimization, master's thesis, National University of Singapore, 2007.
- [30] T. Zhou, D. Tao, and X. Wu, NESVM: a Fast Gradient Method for Support Vector Machines, IEEE International Conference on Data Mining (ICDM), 2010.
- [31] G. Zhou, K. C. Toh, and J. Sun, Efficient algorithms for the smallest enclosing ball problem, *Comput. Optim. Appl.* **30** (2005), 147–160.