SMOOTHING ALGORITHMS FOR COMPUTING THE PROJECTION
ONTO A MINKOWSKI SUM OF CONVEX SETS

Xiaolong Qin\footnote{1Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (email: qxllxajh@163.com)}Nguyen Thai An\footnote{2Institute of Research and Development, Duy Tan University, Danang, Vietnam (email: thai-ian2784@gmail.com)}

Abstract. In this paper, the problem of computing the projection, and therefore the minimum distance, from a point onto a Minkowski sum of general convex sets is studied. Our approach is based on the minimum norm duality theorem originally stated by Nirenberg and the Nesterov smoothing techniques. It is shown that projection points onto a Minkowski sum of sets can be represented as the sum of points on constituent sets so that, at these points, all of the sets share the same normal vector which is the negative of the dual solution. The proposed NESMINO algorithm improves the theoretical bound on number of iterations from \(O(\frac{1}{\epsilon})\) by Gilbert [SIAM J. Contr., vol. 4, pp. 61–80, 1966] to \(O\left(\frac{1}{\sqrt{\epsilon}} \ln(\frac{1}{\epsilon})\right)\), where \(\epsilon\) is the desired accuracy for the objective function. Moreover, the algorithm also provides points on each component sets such that their sum is equal to the projection point.

Keywords: Minimum norm problem, projection onto a Minkowski sum of sets, Nesterov’s smoothing technique, fast gradient method

Mathematical Subject Classification 2000: Primary 49J52, 49M29, Secondary 90C30.

1 Introduction

Let \(A\) and \(B\) be two subsets in \(\mathbb{R}^n\). Recall that the Minkowski sum of these sets is defined by
\[
A + B := \{a + b : a \in A, b \in B\}.
\]
The case of more than two sets is defined in the same way by induction. Note that if all the sets \(A_i\) for \(i = 1, \ldots, m\) are convex, then every linear combination of these sets, \(\sum_{i=1}^{m} \lambda_i A_i\) with \(\lambda_i \in \mathbb{R}\) for \(i = 1, \ldots, m\), is also convex. The Euclidean distance function associated with a subset \(Q\) is defined by
\[
d(x; Q) := \inf\{\|q - x\| : q \in Q\},
\]
where \(\|\cdot\|\) is the Euclidean norm. The optimization problem we are concerned with in this paper is the following minimum norm problem
\[
d\left(0; \sum_{i=1}^{p} T_i(\Omega_i)\right) := \min \left\{\|x\| : x \in \sum_{i=1}^{p} T_i(\Omega_i)\right\}, \tag{1.1}
\]
where \( \Omega_i, \) for \( i = 1, \ldots, p, \) are nonempty convex compact sets in \( \mathbb{R}^n \) and \( T_i : \mathbb{R}^m \to \mathbb{R}^n \) are affine mappings satisfying \( T_i(x) = A_i x + a_i, \) where \( A_i, \) for \( i = 1, \ldots, p, \) are \( n \times m \) matrices and \( a_i, \) for \( i = 1, \ldots, p, \) are given points in \( \mathbb{R}^m. \) Since \( \sum_{i=1}^{p} T_i(\Omega_i) \) is closed and convex, and the norm under consideration is Euclidean, (1.1) has a unique solution, which is the projection from the origin onto \( \sum_{i=1}^{p} T_i(\Omega_i). \) We denote this solution by \( x^* \) throughout this paper.

We assume in problem (1.1) that each constituent set \( \Omega_i \) is simple enough so that the corresponding projection operator \( P_{\Omega_i} \) is easy to compute. It is worth noting that there hasn’t been an algorithm for finding the Minkowski sum of general convex sets except for the cases of balls and polytopes. Moreover, in general, the projection onto a Minkowski sum of sets cannot be represented as the sum of the projections onto constituent sets.

Minimum norm problems for the case of polytopes have been well studied in the literature from both theoretical and numerical point of view; see e.g., [21, 30, 13] and the references therein. The most suitable algorithm for solving (1.1) is perhaps the one suggested by Gilbert [10]. The original Gilbert’s algorithm was devised for solving the minimum norm problem associated with just one convex compact set. The algorithm does not require the explicit projection operator of the given set. Instead, it requires in each step the computation of the support point of the set along with a certain direction. By observation that for a given direction, support point of a Minkowski sum of sets can be represented in term of support points of constituent sets, Gilbert’s algorithm thus can be applied for general case of (1.1). Following [10], Gilbert’s algorithm is a descent method that generates a sequence \( \{ z_k \} \) satisfying \( \| z_k \| \) converges downward to \( \| x^* \| \) within \( O(\frac{1}{k}) \) iterations.

Another effective algorithm for distance computation between two convex objects is the GJK algorithm proposed by Gilbert, Johnson and Keerthi [11] and its enhancing versions [12, 3, 1]. The original GJK algorithm was just restricted to compute the distance between objects which can be approximately represented as convex polytopes. In order to reduce the error of the polytope approximations in finding the minimum distance, Gilbert and Fo [12] modified the original GJK to handle general convex objects. The new modified algorithm is based on Gilbert’s algorithm and has the same bound on number of iterations. It has been observed by many authors that, the algorithm often makes rapid movements toward the solution at its starting iterations, however on many problems, the algorithm turns to stuck when it approaches the final solution; see [16, 20].

When deal with problem (1.1), we are interested in the following questions:

**Question 1:** Is it possible to characterize points on each of constituent sets \( \Omega_i \) so that the sum of their images under corresponding affine mappings is equal to \( x^*. \)

**Question 2:** Is there an alternative algorithm that improves the theoretical complexity bound of Gilbert’s algorithm for solving (1.1)?

In this work, we first use the minimum norm duality theorem originally stated by Nirenberg [27] to establish the Fenchel dual problem for (1.1). From this duality result, we show that each projection onto a Minkowski sum of sets can be represented as the sum of points on constituent sets so that at these points, all the sets share the same normal
vector. For numerically solving the problem, we utilize the smoothing technique developed by Nesterov [24, 25]. To this end, we first approximate the dual objective function by a smooth and strongly convex function and solve this dual problem via a fast gradient scheme. After that we show how an approximate solution for the primal problem, i.e., an approximation for the projection of the origin, can be reconstructed from the dual iterative sequence. Our algorithm improves the theoretical bound on number of iterations from $O\left(\frac{1}{\epsilon} \ln(\frac{1}{\epsilon})\right)$ of the Gilbert algorithm to $O\left(\frac{1}{\sqrt{\epsilon}} \ln(\frac{1}{\epsilon})\right)$. Moreover, the algorithm also provides elements on each constituent sets such that the sum of their images under corresponding linear mappings is equal to the projection $x^\star$.

The rest of the paper is organized as follows. In section 2, we provide tools of convex analysis that are widely used in the sequel. The Nesterov’s smoothing technique and fast gradient method are recalled in section 3. In section 4, we state some duality results concerning the minimum norm problems and give the answer for Question 1. Section 5 is devoted to an overview of Gilbert’s algorithm. Section 6 is the main part of the paper devoted to develop a smoothing algorithms for solving (1.1). Some illustrative examples are provided in section 7.

2 Tools of Convex Analysis

In $n$-dimensional Euclidean space $\mathbb{R}^n$, we use $\langle \cdot, \cdot \rangle$ to denote the inner product, and $\| \cdot \|$ to denote the associated Euclidean norm. An extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be convex if $f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y)$, for all $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. We say that $f$ is strongly convex with modulus $\gamma$ if $f - \frac{\gamma}{2} \| \cdot \|^2$ is a convex function. Let $Q$ be a subset of $\mathbb{R}^n$, the support function of $Q$ is defined by

$$
\sigma_Q(u) := \sup \{ \langle u, x \rangle : x \in Q \}, \ u \in \mathbb{R}^n. \tag{2.2}
$$

It follows directly from the definition that $\sigma_Q$ is positive homogeneous and subadditive. We denote by $S_Q(u)$ the solution set of (2.2). The set-valued mapping $S_Q : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called the support point mapping of $Q$. If $Q$ is compact, then $\sigma_Q(u)$ is finite and $S_Q(u) \neq \emptyset$. Moreover, an element $s_Q(u) \in S_Q(u)$ is a point in $Q$ that is farthest in the direction $u$. Thus $s_Q(u)$ satisfies

$$
s_Q(u) \in Q \text{ and } \sigma_Q(u) = \langle u, s_Q(u) \rangle. \tag{2.3}
$$

In order to study minimum norm problem in which the Euclidean distance is replaced by distances generated by different norms, we consider a more general setting. Let $F$ be a closed, bounded and convex set of $\mathbb{R}^n$ that contains the origin as an interior point. The minimal time function associated with the dynamic set $F$ and the target set $Q$ is defined by

$$
T_F(x; Q) := \inf \{ t \geq 0 : (x + tF) \cap Q \neq \emptyset \}. \tag{2.4}
$$

The minimal time function (2.3) can be expressed as

$$
T_F(x; Q) = \inf \{ \rho_F(\omega - x) : \omega \in Q \}, \tag{2.4}
$$

In order to study minimum norm problem in which the Euclidean distance is replaced by distances generated by different norms, we consider a more general setting. Let $F$ be a closed, bounded and convex set of $\mathbb{R}^n$ that contains the origin as an interior point. The minimal time function associated with the dynamic set $F$ and the target set $Q$ is defined by

$$
T_F(x; Q) := \inf \{ t \geq 0 : (x + tF) \cap Q \neq \emptyset \}. \tag{2.3}
$$

The minimal time function (2.3) can be expressed as

$$
T_F(x; Q) = \inf \{ \rho_F(\omega - x) : \omega \in Q \}, \tag{2.4}
$$
where \( \rho_F(x) := \inf \{ t \geq 0 : x \in tF \} \) is the Minkowski function associated with \( F \). Moreover, \( T_F(\cdot, Q) \) is convex if and only if \( Q \) is convex; see [22]. We denote by

\[
\Pi_F(x; Q) := \{ q \in Q : \rho_F(q - x) = T_F(x; Q) \}
\]

the set of generalized projection from \( x \) to \( Q \).

Note that, if \( F \) is the closed unit ball generated by some norm \( \| \cdot \| \) on \( \mathbb{R}^n \), then we have \( \rho_F = \| \cdot \| \), \( \sigma_F = \| \cdot \|_2 \) and \( T_F(\cdot, Q) \) reduces to the ordinary distance function

\[
d(x; Q) = \inf \{ \| \omega - x \| : \omega \in Q \}, \quad x \in \mathbb{R}^n.
\]

The set \( \Pi_F(x; Q) \) in this case is denoted by \( \Pi(x; Q) := \{ q \in Q : d(x; Q) = \| q - x \| \} \). When \( \| \cdot \| \) is Euclidean norm, we simply use the notation \( P_{\Omega}(x) \) instead. If \( \Omega \) is a nonempty closed convex set, then the Euclidean projection \( P_{\Omega}(x) \) is a singleton for every \( x \in \mathbb{R}^n \).

The following results whose proof can be found in [14] allow us to represent support functions of general sets in term of the support functions of one or more simpler sets.

**Lemma 2.1** Consider the support function \( \sigma \) \text{[22]} . Let \( \Omega, \Omega_1, \Omega_2 \) be subsets of \( \mathbb{R}^m \) and \( T : \mathbb{R}^m \rightarrow \mathbb{R}^n \) satisfying \( T(x) = Ax + a \) be an affine transformation, where \( A \) is an \( n \times m \) matrix and \( a \in \mathbb{R}^n \). The following assertions hold:

(i) \( \sigma_{\Omega} = \sigma_{A \Omega} = \sigma_{\Omega A} \).

(ii) \( \sigma_{\Omega_1 + \Omega_2}(u) = \sigma_{\Omega_1}(u) + \sigma_{\Omega_2}(u) \) and \( \sigma_{\Omega_1 - \Omega_2}(u) = \sigma_{\Omega_1}(u) + \sigma_{\Omega_2}(-u) \), for all \( u \in \mathbb{R}^m \).

(iii) \( \sigma_{T(\Omega)}(v) = \sigma_{\Omega}(A^T v) + \langle v, a \rangle \), for all \( v \in \mathbb{R}^n \).

From Lemma 2.1, we have the following properties for support point mappings.

**Lemma 2.2** Let \( \Omega, \Omega_1, \Omega_2 \) be convex compact subsets of \( \mathbb{R}^m \) and \( T : \mathbb{R}^m \rightarrow \mathbb{R}^n \) be an affine transformation satisfying \( T(x) = Ax + a \), where \( A \) is an \( n \times m \) matrix and \( a \in \mathbb{R}^n \). The following assertions hold:

(i) \( S_{\Omega_1 + \Omega_2}(u) = S_{\Omega_1}(u) + S_{\Omega_2}(u) \), for all \( u \in \mathbb{R}^m \).

(ii) \( S_{T(\Omega)}(v) = T(S_{\Omega}(A^T v)) = A(S_{\Omega}(A^T v)) + a \), for all \( v \in \mathbb{R}^n \).

(iii) If suppose further that \( \Omega \) is a strictly convex set, then \( S_{\Omega}(u) \) is a singleton for any \( u \in \mathbb{R}^m \setminus \{0\} \).

**Proof.** (i) The assumption on the compactness ensures the nonemptyness of involving support points sets. Let any support point \( \bar{w} \in S_{\Omega_1 + \Omega_2}(u) \). This means that \( \bar{w} \in \Omega_1 + \Omega_2 \) and \( \langle u, \bar{w} \rangle = \sigma_{\Omega_1 + \Omega_2}(u) \). There exists \( \bar{w}_1 \in \Omega_1 \) and \( \bar{w}_2 \in \Omega_2 \) such that \( \bar{w} = \bar{w}_1 + \bar{w}_2 \). Employing Lemma 2.1(ii), we have

\[
\langle u, \bar{w}_1 \rangle + \langle u, \bar{w}_2 \rangle = \sigma_{\Omega_1}(u) + \sigma_{\Omega_2}(u).
\]

From the definition of support functions, \( \langle u, \bar{w}_1 \rangle \leq \sigma_{\Omega_1}(u) \) and \( \langle u, \bar{w}_2 \rangle \leq \sigma_{\Omega_2}(u) \). Therefore, equality \( 2.5 \) holds if and only if \( \langle u, \bar{w}_1 \rangle = \sigma_{\Omega_1}(u) \) and \( \langle u, \bar{w}_2 \rangle = \sigma_{\Omega_2}(u) \). Thus, \( \bar{w} \in S_{\Omega_1}(u) + S_{\Omega_2}(u) \) and we have justified the “⊂” inclusion in (i). The converse implication is straightforward.
Now let \( \bar{w} \in \Omega \) such that \( A\bar{w} + a \in S_{T(\Omega)}(u) \). By Lemma 2.1(iii), we have \( \langle v, A\bar{w} + a \rangle = \sigma_{T(\Omega)}(v) = \sigma_{\Omega}(A^Tv) + \langle v, a \rangle \). This is equivalent to \( \langle A^Tv, \bar{w} \rangle = \sigma_{\Omega}(A^Tv) \). Therefore, \( \bar{w} \in S_{\Omega}(A^Tv) \) and thus \( A\bar{w} + a \in A(S_{\Omega}(A^Tv)) + a \). The converse implication of (ii) is proved similarly.

For (iii), suppose that there exist \( \bar{w}_1, \bar{w}_2 \in \Omega \) with \( \bar{w} \neq \bar{w}_2 \) such that \( \langle u, \bar{w}_1 \rangle = \langle u, \bar{w}_2 \rangle = \sigma_{\Omega}(u) \). It then follows from the properties of support function that

\[
\langle u, \frac{\bar{w}_1 + \bar{w}_2}{2} \rangle = \frac{1}{2} \langle u, \bar{w}_1 \rangle + \frac{1}{2} \langle u, \bar{w}_2 \rangle = \frac{1}{2} \sigma_{\Omega}(u) + \frac{1}{2} \sigma_{\Omega}(u) = \sigma_{\Omega}(u).
\]

Since \( \Omega \) is strictly convex, \( \bar{w} := \frac{\bar{w}_1 + \bar{w}_2}{2} \in \text{int}(\Omega) \). Take \( \epsilon > 0 \) small enough such that \( B(\bar{w}; \epsilon) \subset \Omega \). Then the element \( \hat{w} := \bar{w} + \frac{\epsilon}{\|u\|} \in B(\bar{w}; \epsilon) \subset \Omega \). Moreover, since \( u \neq 0 \), we have

\[
\langle u, \hat{w} \rangle = \langle u, \bar{w} \rangle + \frac{\epsilon}{\|u\|} \|u\| = \sigma_{\Omega}(u).
\]

This is a contradiction. The proof is complete. \( \square \)

The **Fenchel conjugate** of a convex function \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) is defined by

\[
f^*(v) := \sup \{ \langle v, x \rangle - f(x) : x \in \mathbb{R}^n \}, \quad v \in \mathbb{R}^n.
\]

If \( f \) is proper and lower semicontinuous, then \( f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) is also a proper, lower semicontinuous convex function. From the definition, support function \( \sigma_Q \) is the Fenchel conjugate of the **indicator function** \( \delta_Q \) of \( Q \) which is defined by \( \delta_Q(x) = 0 \) if \( x \in Q \) and \( \delta_Q(x) = +\infty \) otherwise.

The polar of a subset \( E \subset \mathbb{R}^n \) is the set \( E^\circ = \{ u \in \mathbb{R}^n : \sigma_E(u) \leq 1 \} \). When \( E \) is the closed unit ball of a norm \( \| \cdot \| \), then \( E^\circ \) is the closed unit ball of the corresponding dual norm \( \| \cdot \|_\ast \). Some basis properties of the polar set are collected in the following result whose proof can be found in [29] Proposition 1.23.

**Proposition 2.3** The following assertions hold:

(i) For any subset \( E \), the polar \( E^\circ \) is a closed convex set containing the origin and \( E \subset E^\circ \);

(ii) \( 0 \in \text{int}(E) \) if and only if \( E^\circ \) is bounded;

(iii) \( E = E^{\circ\circ} \) if \( E \) is closed convex and contains the origin;

(iv) If \( E \) is closed convex and contains the origin, then \( \rho_E = \sigma_{E^\circ} \) and \( (\rho_E)^* = \delta_{E^\circ} \).

Thus, if \( F \) is a closed convex and bounded set with \( 0 \in \text{int}(F) \) then \( F^\circ \) is also a closed convex and bounded set with \( 0 \in \text{int}(F^\circ) \). Moreover, from the subadditive property, \( \rho_F = \sigma_{F^\circ} \) is a Lipschitz function with modulus \( \|F^\circ\| := \sup \{ \|x\| : x \in F^\circ \} \).

Let us recall below the Fenchel duality theorem which plays an important role in the sequel. We denote the set of points where a function \( g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) is finite and continuous by \( \text{cont} \).

**Theorem 2.4** (See [2] Theorem 3.3.5) **Given functions** \( f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\} \) and \( g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \), and a linear mapping \( A : \mathbb{R}^m \to \mathbb{R}^n \), the weak duality inequality

\[
\inf_{x \in \mathbb{R}^m} \{ f(x) + g(Ax) \} \geq \sup_{u \in \mathbb{R}^n} \{ -f^*(A^*u) - g^*(-u) \}
\]
holds. If furthermore \( f \) and \( g \) are convex and satisfy the following condition

\[
A(\text{dom } f) \cap \text{cont } g \neq \emptyset,
\]

then the equality holds and the supremum is attained if it is finite.

3 Nesterov’s Smoothing Technique and Fast Gradient Method

In a celebrated work [26], Nesterov introduced a fast first-order method for solving convex smooth problems in which the objective functions have Lipschitz continuous gradients. In contrast to the complexity bound of \( O(1/\epsilon) \) possessed by the classical gradient descent method, Nesterov’s method gives a complexity bound of \( O(1/\sqrt{\epsilon}) \), where \( \epsilon \) is the desired accuracy for the objective function.

When the problem under consideration is nonsmooth in which the objective function has an explicit max-structure as follows

\[
f(u) := \max \{ \langle Au, x \rangle - \phi(x) : x \in Q \}, \quad u \in \mathbb{R}^n,
\]

where \( A \) is an \( m \times n \) matrix and \( \phi \) is a continuous convex function on a compact set \( Q \) of \( \mathbb{R}^m \), in order to overcome the complexity bound \( O(1/\epsilon) \) of the subgradient method, Nesterov [24] made use of the special structure of \( f \) to approximate it by a function with Lipschitz continuous gradient and then applied a fast gradient method to minimize the smooth approximation. With this combination, we can solve the original non-smooth problem up to accuracy \( \epsilon \) in \( O(1/\epsilon) \) iterations. To this end, let \( d \) be a continuous strongly convex function on \( Q \). Let \( \mu \) be a positive number called a smooth parameter. Define

\[
f_\mu(u) := \max \{ \langle Au, x \rangle - \phi(x) - \mu d(x) : x \in Q \}.
\]

Since \( d(x) \) is strongly convex, problem (3.8) has a unique solution. The following statement is a simplified version of [24, Theorem 1].

**Theorem 3.1** (See [24, Theorem 1]) The function \( f_\mu \) in (3.8) is well defined and continuously differentiable on \( \mathbb{R}^n \). The gradient of the function is

\[
\nabla f_\mu(u) = A^\top x_\mu(u),
\]

where \( x_\mu(u) \) is the unique element of \( Q \) such that the maximum in (3.8) is attained. Moreover, \( \nabla f_\mu \) is a Lipschitz function with the Lipschitz constant \( \ell_\mu = \frac{1}{\mu \sigma_1} \| A \|^2 \), and

\[
f_\mu(u) \leq f(u) \leq f_\mu(u) + \mu D \quad \forall u \in \mathbb{R}^n,
\]

where \( D := \max \{ d(x) : x \in Q \} \).

For the reader’s convenience, we conclude this section with a presentation of the simplest optimal method for minimizing smooth strongly convex functions; see [25] and the references therein. Let \( g : \mathbb{R}^n \to \mathbb{R} \) be strongly convex with parameter \( \gamma > 0 \) and its gradient be Lipschitz continuous with constant \( L > \gamma \). Consider problem \( g^* = \inf \{ g(u) : u \in \mathbb{R}^n \} \) and denote by \( u^* \) its unique optimal solution.
### Fast Gradient Method

**INITIALIZE:** $\gamma, v_0 = u_0 \in \mathbb{R}^n$.

Set $k = 0$.

**Repeat the following**

Set $u_{k+1} := v_k - \frac{1}{L} \nabla g(v_k)$

Set $v_{k+1} := u_{k+1} + \frac{\sqrt{L} - \sqrt{\gamma}}{\sqrt{L} + \sqrt{\gamma}} (u_{k+1} - u_k)$

Set $k := k + 1$

Until a stopping criterion is satisfied.

---

By taking into account [25, Theorem 2.2.3], $\{u_k\}_{k=0}^\infty$ satisfies

$$g(u_k) - g^* \leq \left( g(u_0) - g^* + \frac{\gamma}{2} \|u_0 - u^*\|^2 \right) \left( 1 - \sqrt{\frac{2}{L}} \right)^k$$

$$\leq \left( g(u_0) - g^* + \frac{\gamma}{2} \|u_0 - u^*\|^2 \right) e^{-k \sqrt{\frac{\gamma}{L}}}$$

$$\leq 2 (g(u_0) - g^*) e^{-k \sqrt{\frac{\gamma}{L}}},$$

(3.9)

while the last inequality is a consequence of [25, Theorem 2.2.3].

Since $g$ is a differentiable strongly convex function and $u^*$ is its unique minimizer on $\mathbb{R}^n$, we have $\nabla g(u^*) = 0$. Using [25, Theorem 2.1.5], we find

$$\frac{1}{2L} \|\nabla g(u_k)\|^2 \leq g(u_k) - g^* \overset{3.9}{\leq} 2 (g(u_0) - g^*) e^{-k \sqrt{\frac{\gamma}{L}}}.$$  

(3.10)

---

### 4 Duality for Minimum Norm Problems

In this section, we are in a position to give some duality results concerning minimum norm problem (1.1). Let us first recall the duality theorem originally stated by Nirenberg [27].

**Theorem 4.1** (Minimum norm duality theorem) Given $\bar{x} \in \mathbb{R}^n$ and let $d(\cdot; \Omega)$ be the distance function to a nonempty closed convex set $\Omega$ associated with some norm $\| \cdot \|$ on $\mathbb{R}^n$. Then

$$d(\bar{x}; \Omega) = \max \{ \langle u, \bar{x} \rangle - \sigma_{\Omega}(u) : \|u\| \leq 1 \},$$

where the maximum on the right is achieved at some $\bar{u}$. Moreover, if $\bar{w} \in \Pi(\bar{x}; \Omega)$, then $-\bar{u}$ is aligned with $\bar{w} - \bar{x}$, i.e., $\langle -\bar{u}, \bar{w} - \bar{x} \rangle = \|\bar{u}\| \|\bar{w} - \bar{x}\|$.

According to this theorem, the minimum distance from a point to a convex set is equal to the maximum of the distance from the point to hyperplanes separating the point and the set; see Figure 4. A standard proof of this theorem can be found in [19, p. 136]. We also refer the readers to the recent paper [7] for more types of minimum norm duality theorems concerning the width and the length of symmetrical convex bodies.
Lemma 4.2 Let $Q$ be a nonempty closed subset of $\mathbb{R}^n$. Then the generalized projection $\Pi_F(x; Q)$ is nonempty for any $x \in \mathbb{R}^n$.

Proof. From the assumption that $F$ is a closed bounded and convex set that contains the origin as an interior point, $0 \leq T_F(x; Q) < +\infty$ for all $x \in \mathbb{R}^n$ and the following number exists

$$R = \sup \{ r : B(0; r) \subset F^c \} < +\infty.$$ 

Then we have $\rho_F(x) = \sigma_{F^c}(x) \geq R\|x\|$ for all $x \in \mathbb{R}^n$. Fix $x \in \mathbb{R}^n$. For each $n \in \mathbb{N}$, from (2.4) there exists $w_n \in Q$, such that

$$T_F(x; Q) \leq \rho_F(w_n - x) < T_F(x; Q) + \frac{1}{n}. \tag{4.11}$$

It follows from (4.11) and triangle inequality that that

$$R\|w_n\| \leq R (\|w_n - x\| + \|x\|) \leq \rho_F(w_n - x) + R\|x\| \leq T_F(x; Q) + 1 + R\|x\|$$

for all $n$. Thus the sequence $\{w_n\}$ is bounded. We can take a subsequence $\{w_{k_n}\}$ that converges to a point $\bar{w} \in Q$ due to the closedness of $Q$. By taking the limit both sides of (4.11) and using the continuity of $T_F(\cdot; Q)$ and the Minkowski function, we can conclude that $\bar{w} \in \Pi_F(x; Q)$. \hfill \Box

Theorem 4.1 is in fact a direct consequence of the Fenchel duality theorem which is used to prove the following extension for minimal time functions.

Theorem 4.3 The generalized distance $T_F(0; A(\Omega))$ from the origin $0_{\mathbb{R}^n}$ to the image $A(\Omega)$ of a nonempty closed convex set $\Omega \subset \mathbb{R}^m$ under a linear mapping $A : \mathbb{R}^m \to \mathbb{R}^n$ can be computed by

$$T_F(0; A(\Omega)) := \inf \{\rho_F(Aw) : w \in \Omega\} = \max\{-\sigma_{\Omega}(-A^Tu) : u \in F^c\},$$
where the maximum on the right is achieved at some \( \bar{u} \in F^\circ \). If \( A \bar{w} \in \Pi_F (0; A(\Omega)) \) is a projection from the origin to \( A(\Omega) \), then

\[
\langle A\bar{w}, \bar{u} \rangle = \sigma_{F^\circ}(A\bar{w}) = -\sigma_{\Omega}(-A^\top \bar{u}).
\]

**Proof.** Applying Theorem 2.4 for \( g = \rho_F \) and \( f = \delta_{\Omega} \), the following qualification condition holds

\[
\text{Ad} \cap \text{cont}(g) = A(\Omega) \cap \mathbb{R}^n = A(\Omega) \neq \emptyset,
\]

where \( \text{cont}(g) = \mathbb{R}^n \) is due to the fact that \( \rho_F \) is a continuous function on \( \mathbb{R}^n \). It follows that

\[
T_F (0; A(\Omega)) = \min \{ \delta_{\Omega}(x) + \rho_F (Ax) : x \in \mathbb{R}^n \}
\]

\[
= \sup \{ - (\delta_{\Omega})^* (A^\top u) - (\rho_F)^* (-u) : u \in \mathbb{R}^n \}
\]

\[
= \sup \{ -\sigma_{\Omega} (A^\top u) - \delta_{F^\circ} (-u) : u \in \mathbb{R}^n \}
\]

\[
= \sup \{ -\sigma_{\Omega} (-A^\top u) - \delta_{F^\circ} (u) : u \in \mathbb{R}^n \}
\]

\[
= \sup \{ -\sigma_{\Omega} (-A^\top u) : u \in F^\circ \},
\]

and the supremum is attained because \( T_F (0; A(\Omega)) \) is finite. If the supremum on the right is achieved at some \( \bar{u} \in F^\circ \) and the infimum on the left is achieved at some \( \bar{w} \in \Omega \), then

\[
\sigma_{F^\circ} (A\bar{w}) = \rho_F (A\bar{w}) = T_F (0, A(\Omega)) = \max \{ -\sigma_{\Omega} (-A^\top u) : u \in F^\circ \} = -\sigma_{\Omega} (-A^\top \bar{u}).
\]

Since \( \bar{w} \in \Omega \), we also have

\[
\langle -A^\top \bar{u}, \bar{w} \rangle \leq \sigma_{\Omega} (-A^\top u) = -\sigma_{F^\circ} (A\bar{w}).
\]

This implies that \( \langle A^\top \bar{u}, \bar{w} \rangle \geq \sigma_{F^\circ} (A\bar{w}) \). On the other hand, \( \sigma_{F^\circ} (A\bar{w}) \geq \langle A\bar{w}, \bar{u} \rangle = \langle A^\top \bar{u}, \bar{w} \rangle \), because \( \bar{u} \in F^\circ \). Thus, \( \langle A\bar{w}, \bar{u} \rangle = \sigma_{F^\circ} (A\bar{w}) \). This completes the proof. \( \square \)

Note that, given a closed set \( \Omega \), the set \( A(\Omega) \) need not to be closed and therefore, we can not use the min to replace the inf in the primal problem in Theorem 4.3.

**Proposition 4.4** Let \( Q \) be a nonempty, closed convex subset of \( \mathbb{R}^n \). The following holds

\[
T_F (0; Q) := \min \{ \rho_F (q) : q \in Q \} = \max \{ -\sigma_Q (-u) : u \in F^\circ \}. \tag{4.12}
\]

If the maximum on the right is achieved at \( \bar{u} \in F^\circ \) and the infimum on the left is attained at \( \bar{q} \in Q \), then

\[
\langle \bar{q}, \bar{u} \rangle = \sigma_{F^\circ} (\bar{q}) = -\sigma_Q (-\bar{u}). \tag{4.13}
\]

If \( F = \mathbb{B} \) is the Euclidean closed unit ball, then the projection \( \bar{q} \) exists uniquely and

\[
d(0; Q) := \min \{ \| q \| : q \in Q \} = \max \{ -\sigma_Q (-u) : u \in \mathbb{B} \}.
\]

If suppose further that \( 0 \notin Q \), then \( \frac{\bar{q}}{\| \bar{q} \|} \) is the unique solution of the dual problem.
Proof. The first assertion is a direct consequence of Theorem 4.3 with $\Omega = Q$ and $A$ is the identity mapping of $\mathbb{R}^n$. Note that, by Lemma 4.2, the infimum is also attained here. When $F$ is the Euclidean ball, the minimal time function reduces to the Euclidean distance function and therefore the projection $\bar{q} = P_Q(0)$ exists uniquely. If $0 \notin Q$, then $\bar{q} \neq 0$. Moreover, we have $\langle -\bar{q}, x - \bar{q} \rangle \leq 0$ for all $x \in Q$. This implies,
\[ \langle -\frac{\bar{q}}{\|\bar{q}\|}, x \rangle \leq -\|\bar{q}\|, \text{ for all } x \in Q. \]
Hence $\sigma_Q \left( -\frac{\bar{q}}{\|\bar{q}\|} \right) \leq -\|\bar{q}\| = -d(0; Q)$. This means that $\frac{\bar{q}}{\|\bar{q}\|}$ is a solution of the following dual problem
\[ d(0; Q) = \max \{ -\sigma_Q(-u) : u \in \mathcal{B} \}. \]
From (4.13), any dual solution $\bar{u}$ must satisfy $\bar{u} \in S_{F^\circ}(\bar{q})$. Since $F = \mathcal{B}$, we have $F^\circ = \mathcal{B}$ is a strictly convex set. Thus, by Lemma 2.2(iii), $\bar{u} = \frac{\bar{q}}{\|\bar{q}\|}$ is the unique solution of dual problem. The proof is now complete. 

Figure 2: A minimum norm problem with non-Euclidean distance.

From (4.13), for any primal-dual pair $(\bar{q}, \bar{u})$, we have the following relationship
\[ \bar{u} \in S_{F^\circ}(\bar{q}) \quad \text{and} \quad \bar{q} \in S_Q(-\bar{u}). \quad (4.14) \]
This observation seems to be useful from numerical point of view in the sense that if a dual solution $\bar{u}$ is found exactly, then a primal solution $\bar{q}$ can be obtained by taking a support point in $S_Q(-\bar{u})$. However, for a general convex set $Q$, the set $S_Q(-\bar{u})$ might contain more than one point and there might be some points in this set which is not a desired primal solution. Thus, the above task is possible when $S_Q(-\bar{u})$ is a singleton.

When the distance function under consideration is non-Euclidean, the primal problem may have infinitely many solutions and we may not recover a dual solution from a primal one $\bar{q}$ by setting $\frac{\bar{q}}{\|\bar{q}\|}$ as in the Euclidean case.
Example 4.5 In $\mathbb{R}^2$, consider the problem of finding the projection onto the set $Q = \{ x \in \mathbb{R}^2 : 2 \leq x_1 \leq 5 \text{ and } 1 \leq x_2 \leq 4 \}$ in which the distance function generated by the $\ell_\infty$-norm. In this case, $F = \{ x \in \mathbb{R}^2 : \max\{|x_1|, |x_2|\} \leq 1 \}$ and $F^\circ = \{ x \in \mathbb{R}^2 : |x_1| + |x_2| \leq 1 \}$ and we have $T_F(0; Q) = 2$. The primal problem in (4.12) has the solution set $\Pi_F(0; Q) = \{ x \in \mathbb{R}^2 : x_1 = 2 \text{ and } 1 \leq x_2 \leq 2 \}$ and the corresponding dual problem has a unique solution $\tilde{u} = (1, 0)$. We can see that, for any primal solution $\tilde{q}$, the element $\frac{\tilde{q}}{|\tilde{q}|}$ is not a dual solution; see Figure 2.

We now give a sufficient condition for the uniqueness of solution of primal and dual problems in (4.12). We recall the following definition from [23]. The set $F$ is said to be normally smooth if and only if for every boundary point $\bar{x}$ of $F$, the normal cone of $F$ at $\bar{x}$ defined by $N(\bar{x}; F) := \{ u \in \mathbb{R}^n : \langle u, x - \bar{x} \rangle \leq 0, \forall x \in F \}$ is generated exactly by one vector. That means, there exists $a_\bar{x} \in \mathbb{R}^n$ such that $N(\bar{x}; F) = \text{cone}\{a_\bar{x}\}$. From [23 Proposition 3.3], we have that $F$ is normally smooth if and only if its polar $F^\circ$ is strictly convex.

Proposition 4.6 We have the following:

(i) If $Q$ is a nonempty closed and strictly convex set of $\mathbb{R}^n$, then the generalized projection set $\Pi_F(x; Q)$ is a singleton for all $x \in \mathbb{R}^n$.

(ii) If $F$ is normally smooth, then the dual problem in (4.12) has a unique solution.

Proof. (i) It follows from the definitions of minimal time function and generalized projection that $T_F(x; Q) = T_F(0; Q - \{ x \})$ and $\Pi_F(x; Q) = x + \Pi_F(0; Q - \{ x \})$. It suffices to prove that $\Pi_F(0; Q - \{ x \}) \neq \emptyset$. Let $\tilde{u}$ is a dual solution in (4.12). Since $Q$ is nonempty and closed, the set $\Pi_F(0; Q - \{ x \})$ is nonempty by Lemma 4.2. Suppose that $\Pi_F(0; Q - \{ x \})$ contains two distinct elements $q_1 - x \neq q_2 - x$. Then, by relation (4.14), both $q_1$ and $q_2$ belong to the set $S_Q(-\tilde{u})$. This is a contradiction to Lemma 2.2 by the strictly convexity of $Q$ and justifies (i). The proof of (ii) is similar by using the strictly convexity of $F^\circ$. $
$
Minkowski sum of two closed sets is not necessarily closed. For example, for $Q_1 = \{ x \in \mathbb{R}^2 : x_2 \geq e^{x_1} \}$ and $Q_2 = \{ x \in \mathbb{R}^2 : x_2 = 0 \}$, the sum

$$Q_1 + Q_2 = \{ x \in \mathbb{R}^2 : x_2 > 0 \}$$

is an open set. In what follows, in order to ensure the existence of support point for the Minkowski sum, we assume that all component sets are compact.

We now show that (4.14) can allow us to characterize points on each constituent sets in the Minkowski sum so that their sum is equal to the projection point. The answer for Question 1 in the introduction is stated in the following results; see Figure 3 for an illustration.

Corollary 4.7 Let $\{ Q_i \}_{i=1}^p$ be a finite collection of nonempty convex compact sets in $\mathbb{R}^n$. It holds that

$$T_F \left( 0; \sum_{i=1}^p Q_i \right) = -\min \left\{ \sum_{i=1}^p \sigma_{Q_i}(-u) : u \in F^\circ \right\}.$$
Moreover, if the minimum on the right hand side is attained at \( \bar{u} \in F^\circ \), then any generalized projection \( \bar{q} \) of the origin onto the set \( \sum_{i=1}^{p} Q_i \) satisfies

\[
\bar{q} \in S_{Q_1}(-\bar{u}) + \ldots + S_{Q_p}(-\bar{u}).
\]

Thus, the projection \( \bar{q} \) is the sum of points on component sets such that at these points all the sets have the same normal vector \( -\bar{u} \). If \( F = \mathbb{B} \) is the Euclidean closed unit ball, then the projection \( \bar{q} \) exists uniquely and

\[
d(0; \sum_{i=1}^{p} Q_i) = -\min \left\{ \sum_{i=1}^{p} \sigma_{Q_i}(-u) : u \in \mathbb{B} \right\}.
\]

If in addition, \( 0 \notin \sum_{i=1}^{p} Q_i \) then \( \bar{q} \) is the unique solution of the dual problem and we have

\[
\bar{q} \in S_{Q_1}(-\bar{q}) + \ldots + S_{Q_p}(-\bar{q}).
\]

**Proof.** Let \( Q := \sum_{i=1}^{p} Q_i \). Using Lemma 2.1 and Lemma 2.2 we have

\[
\sigma_Q(-u) = \sigma_{Q_1}(-u) + \ldots + \sigma_{Q_p}(-u) \quad \text{and} \quad S_Q(-u) = S_{Q_1}(-u) + \ldots + S_{Q_p}(-u).
\]

Note that, the support point mapping \( S_Q(u) \) does not depend on the magnitude of \( u \), using Proposition 4.4 and relation (4.14), we clarify the desired conclusion easily.

The problem of finding a pair of closest points, and therefore the Euclidean distance, between two given convex compact sets \( P \) and \( Q \) can be reduced to the minimum problem associated with the Minkowski sum \( Q - P \) by observing that \( d(P, Q) = d(0, Q - P) \). A note here is that although there may be several pairs of closest points, the latter problem always has a unique solution which is the projection from 0 onto \( Q - P \). By noting that \( \sigma_{-P}(-u) = \sigma_P(u) \) and \( S_{-P}(-u) = -S_P(u) \), we have the following result; see Figure 6.
Corollary 4.8 Let \( \{Q_i\}_{i=1}^p \) and \( \{P_j\}_{j=1}^l \) be two finite collection of nonempty convex compact sets in \( \mathbb{R}^n \) and let \( P = \sum_{j=1}^l P_j, \ Q = \sum_{i=1}^p Q_i \). It holds that
\[
d(P, Q) = -\min \left\{ \sum_{i=1}^p \sigma_{Q_i}(-u) + \sum_{j=1}^l \sigma_{P_j}(u) : u \in B \right\}.
\]
Moreover, if \( \bar{q} \) is the projection of the origin onto \( Q := Q - P \), if \( (\bar{a}, \bar{b}) \) is a pair of closest points of \( Q \) and \( P \), then \( \bar{q} = \bar{a} - \bar{b} \) and
\[
\bar{a} \in S_{Q_1}(-\bar{q}) + \ldots + S_{Q_p}(-\bar{q}) \quad \text{and} \quad \bar{b} \in S_{P_1}(\bar{q}) + \ldots + S_{P_l}(\bar{q}).
\]
Thus, \( \bar{a} \) is the sum of points in \( Q_i \) for \( i = 1, \ldots, p \) such that at these points all \( Q_i \) have the same normal vector \( -\bar{q} \) and \( \bar{b} \) is the sum of points in \( P_j \) for \( j = 1, \ldots, l \) such that at these points all \( P_j \) have the same normal vector \( \bar{q} \).

5 The Gilbert Algorithm

We now give an overview and clarify how the Gilbert algorithm can be applied for solving \( (1.1) \). Let us define the function \( g : \mathbb{R}^n \times Q \to \mathbb{R} \) by
\[
g_Q(z, x) := \sigma_Q(z) - \langle z, x \rangle,
\]
where \( Q = \sum_{i=1}^p T_i(\Omega_i) \). From the definition, \( g_Q(-z, z) \geq 0 \) for all \( z \in Q \). A point \( z \in Q \) is the solution of \( (1.1) \) if and only if \( \langle -z, x - z \rangle \leq 0 \) for all \( x \in Q \). This amounts to saying that \( g_Q(-z, z) = 0 \).

Lemma 5.1 If two points \( z \) and \( \bar{z} \) satisfy \( \|\bar{z}\|^2 - \langle z, \bar{z} \rangle > 0 \), then there is a point \( \bar{z} \) in the line segment \( co\{z, \bar{z}\} \) such that \( \|\bar{z}\| < \|z\| \).

Proof. If \( \|\bar{z}\|^2 < \langle z, \bar{z} \rangle \), then we can choose \( \bar{z} = \bar{z} \). Consider the case \( \|\bar{z}\|^2 < \langle z, \bar{z} \rangle \). By combining with the assumption \( \|z\|^2 - \langle z, \bar{z} \rangle > 0 \), we have
\[
0 < \lambda* := \frac{\|z\|^2 - \langle z, \bar{z} \rangle}{\|z - \bar{z}\|^2} < 1.
\]
This implies the quadratic function
\[
f(\lambda) = \|\bar{z} - z\|^2 \lambda^2 + 2\langle z, \bar{z} - z \rangle \lambda + \|z\|^2
\]
attains its minimum on \([0, 1]\) at \( \lambda^* \) and therefore
\[
f(\lambda^*) = \|z + \lambda^*(\bar{z} - z)\|^2 < f(0) = \|z\|^2.
\]
Thus \( \bar{z} := z + \lambda^*(\bar{z} - z) \) is the desired point.

The Gilbert algorithm can be interpreted as follows. Starting from some \( z \in Q \), if \( g_Q(-z, z) = 0 \) then \( z \) is the solution. If \( g_Q(-z, z) > 0 \), then \( \bar{z} \in S_Q(-z) \) satisfies \( \|z\|^2 - \langle z, \bar{z} \rangle > 0 \). Using Lemma 5.1 we find a point \( \bar{z} \) on the line segment connecting \( z \) and \( \bar{z} \) such that \( \|\bar{z}\| < \|z\| \). The algorithm is outlined as follows.
Gilbert’s Algorithm

0. Initialization step: Take arbitrary point $z_0 \in Q$.
1. If $g_Q(z, z) = 0$, then return $z = x^*$ is the solution
else, set $\bar{z} \in S_Q(z)$.
2. Compute $\tilde{z} \in \text{co}\{z, \bar{z}\}$ which has minimum norm, set $z = \tilde{z}$
and go back to step 1.

Figure 4 illustrates some iterations of Gilbert’s algorithm for finding closest point to
an ellipse in two dimension. Lemma 5.1 also suggests an effective way to find $\tilde{z}$ in step 3.
We have $\tilde{z} := z + \lambda^*(\bar{z} - z)$, where

$$
\lambda^* = \begin{cases}
1, & \text{if } \|\bar{z}\|^2 - \langle z, \bar{z} \rangle \\
\|\bar{z} - z\|^2, & \text{otherwise}.
\end{cases}
$$

To implement the algorithm, it remains to show how to compute a supporting point for
$Q = \sum_{i=1}^p T_i(\Omega_i)$. Fortunately, this can be done by using Lemma 2.2.

Gilbert showed that, if $\{z_k\}_{k=1}^\infty$ generated by the algorithm does not stop with $z = x^*$
at step 1 within a finite number of iterations, then $z_k \to x^*$ asymptotically. According to
[10, Theorem 3], we have

$$
\|z_k\| - \|x^*\| \leq \frac{C_1}{k} \quad \text{and} \quad \|z_k - x^*\| \leq \frac{C_2}{\sqrt{k}},
$$

where $C_1$ and $C_2$ are some positive constants. From the above estimates, in order to find
an $\epsilon$ - approximate solution, i.e., a point $z$ such that $\|z\| - \|x^*\| \leq \epsilon$, we need to perform the
algorithm in $O(\frac{1}{\epsilon})$ iterations. Gilbert also showed the bounds (5.16) are sharp in the sense
that within a constant multiplicative factor it is impossible to obtain bounds on $\|z_k\| - \|x^*\|
and $\|z_k - x^*\|$ which approach zero more rapidly than those given (5.16); see [10, Example 1].
6 Smoothing Algorithm for Minimum Norm Problems

Our approach for numerically solving (1.1) is based on the minimum norm duality Theorem 4.3 and the Nesterov smoothing technique [24]. Let us first consider the function of the following type

$$\sigma_{A,Q}(u) = \sup \{ \langle Au, x \rangle : x \in Q \}, \ u \in \mathbb{R}^n,$$

where $A$ is an $m \times n$ matrix and $Q$ is a closed bounded subset of $\mathbb{R}^m$. Observe that $\sigma_{A,Q}(u)$ is the composition of a linear mapping and the support function of $Q$. As we will see, this function can be approximated by the following function

$$\sigma_{A,Q}^\mu(u) = \sup \{ \langle Au, x \rangle - \frac{\mu}{2} \|x\|^2 : x \in Q \}, \ u \in \mathbb{R}^n.$$

The following statement is a directly consequence of Theorem 3.1. However, the approximate function as well as its gradient, in this case, has closed form that is expressed in term of the Euclidean projection. This feature makes it reliable from numerical point of view.

**Proposition 6.1** The function $\sigma_{A,Q}^\mu$ has the following explicit representation

$$\sigma_{A,Q}^\mu(u) = \frac{\|Au\|^2}{2\mu} - \frac{\mu}{2} \left[ d\left( \frac{Au}{\mu} ; Q \right) \right]^2$$

and is continuous differentiable on $\mathbb{R}^n$ with its gradient given by

$$\nabla \sigma_{A,Q}^\mu(u) = A^\top P_Q \left( \frac{Au}{\mu} \right).$$

The gradient $\nabla \sigma_{A,Q}^\mu$ is a Lipschitz function with constant $\ell_\mu = \frac{1}{\mu} \|A\|^2$. Moreover,

$$\sigma_{A,Q}^\mu(u) \leq \sigma_{A,Q}(u) \leq \sigma_{A,Q}^\mu(u) + \frac{\mu}{2} \|Q\|^2 \text{ for all } u \in \mathbb{R}^n, \quad (6.17)$$

where $\|Q\| := \sup\{\|q\| : q \in Q\}$.

**Proof.** We have

$$\sigma_{A,Q}^\mu(u) = \sup \{ \langle Au, x \rangle - \frac{\mu}{2} \|x\|^2 : x \in Q \}$$

$$= \sup \left\{ -\frac{\mu}{2} \left( \|x\|^2 - \frac{2}{\mu} \langle Au, x \rangle \right) : x \in Q \right\}$$

$$= -\frac{\mu}{2} \inf \left\{ \|x - \frac{Au}{\mu}\|^2 - \frac{\|Au\|^2}{\mu^2} : x \in Q \right\}$$

$$= \frac{\|Au\|^2}{2\mu} - \frac{\mu}{2} \inf \left\{ \|x - \frac{Au}{\mu}\|^2 : x \in Q \right\}$$

$$= \frac{\|Au\|^2}{2\mu} - \frac{\mu}{2} \left[ d\left( \frac{Au}{\mu} ; Q \right) \right]^2.$$
Since \( \psi(x) := [d(x; Q)]^2 \) is a differentiable function satisfying \( \nabla \psi(x) = 2[x - P_Q(x)] \) for all \( x \in \mathbb{R}^m \), we find from the chain rule that

\[
\nabla \sigma_{A,Q}^\mu(u) = \frac{1}{\mu} A^\top(Au) - \frac{\mu}{2} \left[ \frac{2}{\mu} A^\top \left( \frac{Au}{\mu} - P_Q(\frac{Au}{\mu}) \right) \right] = A^\top P_Q \left( \frac{Ax}{\mu} \right)
\]

From the property of the projection mapping onto convex sets and Cauchy-Schwarz inequality, we find, for any \( u, v \in \mathbb{R}^n \), that

\[
\| \nabla \sigma_{A,Q}^\mu(u) - \nabla \sigma_{A,Q}^\mu(v) \| = \| A^\top P_Q \left( \frac{Au}{\mu} \right) - A^\top P_Q \left( \frac{Av}{\mu} \right) \|^2 \\
\leq \| A \|^2 \| P_Q \left( \frac{Au}{\mu} \right) - P_Q \left( \frac{Av}{\mu} \right) \|^2 \\
\leq \| A \|^2 \left< \frac{Au - Av}{\mu} , P_Q \left( \frac{Au}{\mu} \right) - P_Q \left( \frac{Av}{\mu} \right) \right> \\
= \frac{\| A \|^2}{\mu} \left< u - v , A^\top P_Q \left( \frac{Au}{\mu} \right) - A^\top P_Q \left( \frac{Av}{\mu} \right) \right> \\
= \frac{\| A \|^2}{\mu} \frac{\| u - v \|}{\| \nabla \sigma_{A,Q}^\mu(u) - \nabla \sigma_{A,Q}^\mu(v) \|}
\]

This implies that

\[
\| \nabla \sigma_{A,Q}^\mu(u) - \nabla \sigma_{A,Q}^\mu(v) \| \leq \frac{\| A \|^2}{\mu} \| u - v \|.
\]

The lower and upper bounds in (6.17) follow from

\[
\langle Au, x \rangle - \frac{\mu}{2} \| x \|^2 \leq \langle Au, x \rangle \leq \langle Au, x \rangle - \frac{\mu}{2} \| x \|^2 + \frac{\mu}{2} \sup \{ \| q \|^2 : q \in Q \},
\]

for all \( x \in Q \). The proof is now complete. \( \square \)

From Proposition 4.4, we have the duality result below

\[
d \left( 0; \sum_{i=1}^p T_i(\Omega_i) \right) = \min \left\{ \sum_{i=1}^p \sigma_{\Omega_i}(-A_i^\top u) - \langle u, \sum_{i=1}^p a_i \rangle : u \in B \right\}.
\]

We now make use of the strong convexity of the squared Euclidean norm to state another dual problem for (1.1) in which the dual objective function is strongly convex.

**Proposition 6.2** The following duality result holds

\[
d^2 \left( 0; \sum_{i=1}^p T_i(\Omega_i) \right) = \min \left\{ \sum_{i=1}^p \sigma_{\Omega_i} (A_i^\top u) + \langle u, \sum_{i=1}^p a_i \rangle + \frac{1}{4} \| u \|^2 : u \in \mathbb{R}^n \right\}. \tag{6.18}
\]
**Proof.** Observe that the Fenchel dual of the function \( \| \cdot \|_2 \) is \( \frac{1}{4} \| \cdot \|_2^2 \). Applying Theorem 2.4 again, for \( Q := \sum_{i=1}^{p} T_i(\Omega_i) \), we have

\[
[d(0; Q)]^2 = \inf \{ \|x\|^2 : x \in Q \}
= \max\{-(\delta_Q)^*(u) - (\| \cdot \|_2^2)^*(-u) : u \in \mathbb{R}^n\}
= \max\{-\sigma_Q(u) - \frac{1}{4}\|u\|^2 : u \in \mathbb{R}^n\}
= -\min\{\sigma_Q(u) + \frac{1}{4}\|u\|^2 : u \in \mathbb{R}^n\}.
\]

The result now follows directly from Lemma 2.1. \( \square \)

In order to solve minimum norm problem (1.1), we solve dual problem (6.18) by approximating the dual objective function by a smooth and strongly convex function with Lipschitz continuous gradient and then apply a fast gradient scheme to this smooth one.

Let us define the dual objective function by

\[
f(u) := \sum_{i=1}^{p} \sigma_{\Omega_i} (A_i^\top u) + \langle u, \sum_{i=1}^{p} a_i \rangle + \frac{1}{4}\|u\|^2, \quad u \in \mathbb{R}^n.
\]

The following result is a direct consequence of Proposition 6.1.

**Proposition 6.3** The function \( f(u) \) has the following smooth approximation

\[
f_\mu(u) := \sum_{i=1}^{p} \left( \frac{\|A_i^\top u\|^2}{2\mu} - \frac{\mu}{2}[d(A_i^\top u; \Omega_i)]^2 \right) + \langle u, \sum_{i=1}^{p} a_i \rangle + \frac{1}{4}\|u\|^2, \quad u \in \mathbb{R}^n.
\]

Moreover, \( f_\mu \) is a strongly convex function with modulus \( \gamma = 2 \) and its gradient is given by

\[
\nabla f_\mu(u) = \sum_{i=1}^{p} A_i P_{\Omega_i} \left( \frac{A_i^\top u}{\mu} \right) + \sum_{i=1}^{p} a_i + \frac{1}{2} u.
\]

The Lipschitz constant of \( \nabla f_\mu \) is

\[
L_\mu := \sum_{i=1}^{p} \frac{\|A_i\|^2}{\mu} + \frac{1}{2}.
\]

Moreover, we have the following estimate

\[
f_\mu(u) \leq f(u) \leq f_\mu(u) + \mu D_f,
\]

where \( D_f := \frac{1}{2} \sum_{i=1}^{p} \|\Omega_i\|^2 < \infty. \)

We now apply the Nesterov fast gradient method introduced in Section 3 for minimizing the smooth and strongly convex function \( f_\mu \). We will show how to recover an approximately optimal solution for primal problem (1.1) from the dual iterative sequence. The NEsterov Smoothing algorithm for MINimum NOrm Problem (NESMINO) is outlined as follows:
NESMINO

**INITIALIZE:** $\Omega_i, A_i, a_i$ for $i = 1, \ldots, p$ and $v_0, u_0, \mu$.

Set $k = 0$.

**Repeat the following**

- Compute $\nabla f_\mu(v_k)$ using (6.19)
- Compute $L_\mu$ using (6.20)
- Set $u_{k+1} := v_k - \frac{1}{L_\mu} \nabla f_\mu(v_k)$
- Set $v_{k+1} := u_{k+1} + \frac{\sqrt{L_\mu} - \sqrt{2}}{\sqrt{L_\mu} + \sqrt{2}} (u_{k+1} - u_k)$
- Set $k := k + 1$

Until a stopping criterion is satisfied.

We denote by $u^*_\mu$ the unique minimizer of $f_\mu$ on $\mathbb{R}^n$. We also denote by $u^*$ a minimizer of $f$ and by $f^* := f(u^*) = \inf_{x \in \mathbb{R}^n} f(x)$ its optimal value on $\mathbb{R}^n$. From the duality result (6.18), we have

$$f^* = -\left[ d \left( 0; \sum_{i=1}^p T_i(\Omega_i) \right) \right]^2.$$  

We say that $x \in \sum_{i=1}^p T_i(\Omega_i)$ is an $\epsilon$-approximate solution of problem (1.1) if it satisfies

$$\|x\| - \|x^*\| \leq \epsilon.$$

**Theorem 6.4** Let $\{u_k\}_{k=1}^\infty$ be the sequence generated by NESMINO algorithm. Then the sequence $\{y_k\}_{k=1}^\infty$ defined by

$$y_k := \sum_{i=1}^p A_i P_{\Omega_i} \left( \frac{A_i^T u_k}{\mu} \right) + a_i$$

converges to an $\epsilon$-approximate solution of minimum norm problem (1.1) within $k = O \left( \frac{1}{\sqrt{\epsilon}} \ln \left( \frac{1}{\epsilon} \right) \right)$ iterations.

**Proof.** Using (3.9), we find that $\{u_k\}_{k=0}^\infty$ satisfies

$$f_\mu(u_k) - f_\mu^* \leq 2 \left( f_\mu(u_0) - f_\mu^* \right) e^{-k\sqrt{\frac{1}{\mu}}}.$$  

(6.22)

From $f_\mu(u_0) \leq f(u_0)$ and the following estimate

$$f_\mu^* = f_\mu(u_\mu^*) \geq f(u_\mu^*) - \mu D_f \geq f(u^*) - \mu D_f = f^* - \mu D_f,$$

we have

$$f_\mu(u_0) - f_\mu^* \leq f(u_0) - f^* + \mu D_f.$$  

(6.23)

Moreover, since $f_\mu(u_k) - f_\mu^* \geq f_{\mu}(u_k) - \mu D_f - f^*$, we find from (6.22) and (6.23) that

$$f(u_k) - f^* \leq \mu D_f + f_\mu(u_k) - f_\mu^*$$

$$\leq \mu D_f + 2 \left( f(u_0) - f^* + \mu D_f \right) e^{-k\sqrt{\frac{1}{\mu}}} \text{ for all } k \geq 0.$$  

(6.24)
Since $f_\mu$ is a differentiable strongly convex function and $u_\mu^*$ is its unique minimizer on $\mathbb{R}^n$, we have $\nabla f_\mu(u_\mu^*) = 0$. It follows from (3.10) that

$$\frac{1}{2L_\mu} \| \nabla f_\mu(u_k) \|^2 \leq f_\mu(u_k) - f_\mu^* \leq 2 \left( f_\mu(u_0) - f_\mu^* \right) e^{-k \sqrt{\frac{\mu}{L_\mu}}}.$$  

This implies

$$\| \nabla f_\mu(u_k) \|^2 \leq 4L_\mu (f_\mu(u_0) - f_\mu^*) e^{-k \sqrt{\frac{\mu}{L_\mu}}} \leq 4L_\mu (f(u_0) - f^*) e^{-k \sqrt{\frac{\mu}{L_\mu}}}. \quad (6.25)$$

For each $k$ and for each $i \in \{1, \ldots, m\}$, let $x_k^i$ be the unique solution to the problem

$$\sigma_{\mu, \Omega_i} \left( A_i^T u_k \right) := \sup \left\{ \langle A_i^T u_k, x \rangle - \frac{\mu}{2} \| x \|^2 : x \in \Omega_i \right\}.$$ 

We have

$$\sup \left\{ \langle A_i^T u_k, x \rangle - \frac{\mu}{2} \| x \|^2 : x \in \Omega_i \right\} = \sup \left\{ \frac{\| A_i^T u_k \|^2}{2\mu} - \frac{\mu}{2} \left\| x - \frac{A_i^T u_k}{\mu} \right\|^2 : x \in \Omega_i \right\}$$

$$= \frac{\| A_i^T u_k \|^2}{2\mu} - \left[ d \left( \frac{A_i^T u_k}{\mu}, \Omega_i \right) \right]^2.$$

Hence $x_k^i = P_{\Omega_i} \left( \frac{A_i^T u_k}{\mu} \right)$. For each $k$, set

$$d_k := \left\| \sum_{i=1}^{p} \left( A_i x_k^i + a_i \right) \right\|^2 - \left[ d \left( 0, \sum_{i=1}^{p} T_i(\Omega_i) \right) \right]^2.$$ 

Observe $y_k := \sum_{i=1}^{p} \left( A_i x_k^i + a_i \right) \in \sum_{i=1}^{p} T_i(\Omega_i)$. From the property of the projection onto convex sets, we have

$$\| y_k - x^* \|^2 = \| y_k \|^2 - \| x^* \|^2 + 2 \langle -x^*, y_k - x^* \rangle \leq \| y_k \|^2 - \| x^* \|^2 = d_k.$$

This implies that $\{ y_k \}$ converges to $x^*$ whenever $d_k \to 0$ as $k \to \infty$. Moreover, we have

$$2\| x^* \| (\| y_k \| - \| x^* \|) \leq (\| y_k \| + \| x^* \|) (\| y_k \| - \| x^* \|) = \| y_k \|^2 - \| x^* \|^2 = d_k. \quad (6.26)$$
We have the following

\[ d_k = \| \sum_{i=1}^{p} (A_i x_k^i + a_i) \|^2 + f^* \]

\[ = \| \sum_{i=1}^{p} (A_i x_k^i + a_i) \|^2 + f_\mu(u_k) + f^* - f_\mu(u_k) \]

\[ = \| \sum_{i=1}^{p} (A_i x_k^i + a_i) \|^2 + \sum_{i=1}^{p} [\langle A_i^T u_k, x_k^i \rangle - \mu \frac{1}{2} \| x_k^i \|^2 + \langle u_k, \sum_{i=1}^{p} a_i \rangle + \frac{1}{4} \| u_k \|^2 \]

\[ + f^* - f_\mu(u_k) \]

\[ = \| \sum_{i=1}^{p} (A_i x_k^i + a_i) \|^2 + \langle u_k, \sum_{i=1}^{p} (A_i x_k^i + a_i) \rangle + \frac{1}{4} \| u_k \|^2 - \mu 2 \sum_{i=1}^{p} \| x_k^i \|^2 \]

\[ + f^* - f_\mu(u_k) \]

\[ = \| \sum_{i=1}^{p} A_i P_O(\frac{A_i^T u_k}{\mu}) + \sum_{i=1}^{p} a_i + \frac{1}{2} \| u_k \|^2 - \mu 2 \sum_{i=1}^{p} \| x_k^i \|^2 + f^* - f_\mu(u_k) \]

\[ = \| \nabla f_\mu(u_k) \|^2 - \mu 2 \sum_{i=1}^{p} \| x_k^i \|^2 + f^* - f_\mu(u_k) \]

Observe \(|f_\mu(u_k) - f^*| \leq 2 \mu D_f \) and \( \sum_{i=1}^{p} \| x_k^i \|^2 \leq 2D_f \). Taking into account (6.24) and (6.25), we have

\[ d_k \leq \| \nabla f_\mu(u_k) \|^2 + |f(u_k) - f^*| + 2 \mu D_f \]

\[ \leq 4L_\mu(f(u_0) - f^*)e^{-k \sqrt{\frac{\mu}{\gamma}}} + \mu D_f \]

\[ + 2(f(u_0) - f^*)e^{-k \sqrt{\frac{\mu}{\gamma}}} + 2 \mu D_f \]

\[ \leq 2(2L_\mu + 1) (f(u_0) - f^* + \mu D_f) e^{-k \sqrt{\frac{\mu}{\gamma}}} + 3 \mu D_f \]

Now, for a fixed \( \epsilon > 0 \), in order to achieve an \( \epsilon \) - approximate solution for the primal problem, we should force each of the two terms in the above estimate less than or equal to \( \frac{\epsilon}{2} \). If we choose the value of smooth parameter \( \mu \) to be \( \frac{\epsilon}{6D_f} \), we have \( d_k \leq \epsilon \) when

\[ k \geq \sqrt{\frac{L_\mu}{\gamma} \ln \left( \frac{4(2L_\mu + 1) (f(u_0) - f^* + \frac{\epsilon}{6})}{\epsilon} \right)} \]

(6.27)

where \( L_\mu = \frac{\sum_{i=1}^{p} 6 |A_i|^2 D_i}{\epsilon} + \frac{1}{2} \).

Thus, from (6.26), we can find an \( \epsilon \) - approximate solution for primal problem within \( k = O \left( \frac{1}{\sqrt{\epsilon}} \ln \left( \frac{1}{\epsilon} \right) \right) \) iterations. The proof is complete.

We highlight the fact that the algorithm does not require computation of the Minkowski sum but rather only the projection onto each of the constituent sets \( \Omega_i \). Fortunately, many
useful projection operators are easy to compute. Explicit formula for projection operator $P_\Omega$ exists when $\Omega$ is a closed Euclidean ball, a closed rectangle, a hyperplane, or a half-space. Although there are no analytic solutions, fast algorithms for computing the projection operators exist for the cases of unit simplex, the closed $\ell_1$ ball (see [8, 5]), or the ellipsoids (see [6]).

In some cases, by making use of the special structure of the support function of $\Omega$, we can have a suitable smoothing technique in order to avoid working with implicit projection operator $P_\Omega$ or to employ some fast projection algorithm. We consider two important cases as follows:

**The case of ellipsoids.** Consider the case of ellipsoids associated with Euclidean norm

$$E(A, c) := \{ x \in \mathbb{R}^n : (x - c)\mathsf{T} A^{-1} (x - c) \leq 1 \},$$

where the shape matrix $A$ is positive definite and the center $c$ is some given point in $\mathbb{R}^n$. It is well known that the support function of this Ellipsoid is $\sigma_E(u) = \sqrt{u^\top A u + u^\top c}$ and the support point in direction $u$ is $s_E(u) = \frac{Au}{\sqrt{u^\top A u}} + c$. We can rewrite the support function as follows

$$\sigma_E(u) = \sigma_B\left(A^{1/2}u\right) + u^\top c,$$

where $B$ stands for the closed unit Euclidean ball and $A^{1/2}$ is the square root of $A$. The smooth approximation $g_\mu$ of function $g = \sigma_E$ has the following explicit representation

$$g_\mu(u) = \frac{\|A^{1/2}u\|^2}{2\mu} - \frac{\mu}{2}\left[d\left(\frac{A^{1/2}u}{\mu}; B\right)\right]^2 + u^\top c.$$

and is differentiable on $\mathbb{R}^n$ with its gradient given by $\nabla g(u) = A^{1/2}P_B\left(\frac{A^{1/2}u}{\mu}\right) + c$. Thus, instead of projecting onto the Ellipsoid, we just need to project onto the closed unit ball.

**The case of polytopes.** Consider the polytope $S = \text{conv}\{a_1, \ldots, a_m\}$ generated by

| Iteration | NESMINO   | Gilbert          |
|-----------|-----------|------------------|
| 1         | (0, 0)    | (1.5, 1.5)      |
| 100       | (0, 1)    | (0.0090, 1.0177) |
| 300       | (0, 1)    | (0.0032, 1.0063) |
| 1000      | (0, 1)    | (0.0001, 1.0002) |
| 10000     | (0, 1)    |                  |

Figure 5: Comparison between NESMINO algorithm and Gilbert’s algorithm for solving a minimum norm problem involving polytopes
m point in \( \mathbb{R}^n \). By [28, Theorem 32.2], we have

\[
\sigma_S(u) = \sup\{ \langle u, x \rangle : x \in S \} = \max_{1 \leq i \leq m} \langle u, a_i \rangle,
\]

and the support point \( s_S(u) \) of \( S \) is some point \( a_i \) such that \( \langle u, a_i \rangle = \sigma_S(u) \). Observe that, for \( \alpha = (\alpha_1, \ldots, \alpha_m) \top \in \mathbb{R}^m \), we have

\[
\max_{1 \leq i \leq m} \alpha_i = \sup \{ x_1 \alpha_1 + \ldots + x_m \alpha_m : x_i \geq 0, \sum_{i=1}^m x_i = 1 \} = \sup \{ \langle \alpha, x \rangle : x \in \Delta_m \}.
\]

Therefore, \( \sigma_S(u) = \max_{1 \leq i \leq m} \langle u, a_i \rangle = \sup \{ \langle Au, x \rangle : x \in \Delta_m \} = \sigma_{\Delta_m}(Au) \), where \( A = \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix} \) and \( \Delta_m \) is the unit simplex in \( \mathbb{R}^m \). The smooth approximate function of \( g = \sigma_S \) is

\[
g_\mu(u) = \frac{\|Au\|^2}{2\mu} - \frac{\mu}{2} \left[ d_{\Delta_m}(Au) \right]^2, \quad \text{with} \quad \nabla g_\mu(u) = A^\top P_{\Delta_m} \left( \frac{Au}{\mu} \right).
\]

We thus can employ the fast and simple algorithms for computing the projection onto a unit simplex, for example in [4], instead of projection onto a polytope.

**Remark 6.5** In NESMINO algorithm, a smaller smooth parameter \( \mu \) is often better because it reduces the error when approximate \( f \) by \( f_\mu \). However, a small \( \mu \) implies a large value of the Lipschitz constant \( L_\mu \) which in turn reduces the convergence rate by (6.27). Thus the time cost of the algorithm is expensive if we fix a value for \( \mu \) ahead of time. In practice, a sequence of smooth problems with decreasing smooth parameter \( \mu \) is solved and the solution of the previous problem is used as the initial point for the next one. The algorithm stops when a preferred \( \mu^* \) is attained. The optimization scheme is outlined as follows.

**INITIALIZE:** \( \Omega, A_i, a_i \) for \( i = 1, \ldots, p \) and \( w_0, \sigma \in (0, 1) \), \( \mu_0 > 0 \) and \( \mu^* > 0 \).

Set \( k = 0 \).

**Repeat the following**

1. Apply NESMINO algorithm with \( \mu = \mu_k, u_0 = v_0 = w_k \) to find

\[
w_{k+1} = \arg\min_{w \in \mathbb{R}^n} f_\mu(w).
\]

2. Update \( \mu_{k+1} := \sigma \mu_k \) and set \( k := k + 1 \).

**Until** \( \mu \leq \mu^* \).

**7 Illustrative Examples**

We now implement NESMINO and Gilbert’s algorithm to solve minimum norm problem (1.1) in a number of examples by MATLAB. We terminate the NESMINO when \( \|\nabla f_\mu(u_k)\| \leq \epsilon \), for some tolerance \( \epsilon > 0 \). In Gilbert’s algorithm, we relax the stopping criterion \( g_\Omega(-z, z) = 0 \) to \( g_\Omega(-z, z) \leq \delta \), for some \( \delta > 0 \). The parameters described in Remark 6.5 are chosen as follows:

\[
\mu_0 = 100, \sigma = 0.1, \mu^* = 10^{-3}, \epsilon = 10^{-3}, w_0 = 0,
\]
and we use $\delta = 10^{-4}$. All the tests are implemented on a personal computer with an Intel Core i5 CPU 1.6 GHz and 4G of RAM. Figures in this section are plotted via the Multi-Parametric Toolbox [IS] and Ellipsoidal Toolbox [LE].

Let us first give a simple example showing that when the sets involved are polytopes, the Gilbert’s algorithm may have zigzag phenomenon and may become very slow as it approaches the final solution.

**Example 7.1** Consider the minimum norm problem associated with a polytope $P$ in $\mathbb{R}^2$ whose vertices given by the columns of the following matrix

\[
\begin{pmatrix}
-2 & 2 & 1 \\
1 & 1 & 2
\end{pmatrix}
\]

The NESMINO algorithm with a fixed value $\mu = 0.1$ converges to the optimal solution $x^* = (0; 1)$ within nearly 100 steps. In contrast, if starting from $z_0 = \left(\frac{3}{2}, \frac{3}{2}\right)$, the approximate values $(z_1; z_2)$ in Gilbert’s algorithm are still changing after $10^4$ iterations. In this case, as the number of iterations is increasing, the Gilbert algorithm alternately chooses the two vertices $(-2, 1)$ and $(2, 1)$ as support points of $P$ and converges slowly to $x^* = (0; 1)$; see Figure 5.

![Figure 5](image)

Figure 5: Minimum distance between ellipse $E_3$ and the sum $E_1 + E_2$ of two other ellipses.

**Example 7.2** In $\mathbb{R}^2$, consider the problem of computing the projection onto the sum of a polytope $P$ whose vertices given by the columns of the following matrix

\[
\begin{pmatrix}
4 & 4 & 2 & 3 \\
2 & 5 & 4 & 1
\end{pmatrix}
\]

and two ellipses $E_1(A_1, c_1)$, $E_2(A_2, c_2)$ with shape matrices and centers respectively given by

\[
A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad c_1 = \begin{pmatrix} 4 \\ -4 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}.
\]
The NESMINO algorithm yields an approximate solution \( x^* = (7.2841, -1.4787) \). The algorithm also gives \( x_1 = (2, 4) \), \( x_2 = (3.0101, -3.8995) \), \( x_3 = (2.2740, -1.5792) \) which respectively belongs to \( P \), \( E_1 \), \( E_2 \) such that \( x^* = x_1 + x_2 + x_3 \). This result is depicted in Figure 8.

**Example 7.3** In this example, we apply NESMINO to find the minimum distance between a Minkowski of two ellipses \( E_1(A_1, c_1), E_2(A_2, c_2) \) with shape matrices and centers respectively given by

\[
A_1 = \begin{pmatrix} 1.5 & -1 \\ -1 & 1.5 \end{pmatrix}, \quad c_1 = \begin{pmatrix} 15 \\ 5 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 10 \\ -5 \end{pmatrix}
\]

and another ellipse \( E_3(A_3, c_3) \) with \( c_3 = \begin{pmatrix} -5 \\ 10 \end{pmatrix}, A_3 = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \). The NESMINO yields the distance \( d = 27.2347 \) that is the norm of the projection \( x^* = (25.4219, -9.7703) \) of the origin onto \( E_1 + E_2 - E_3 \). Moreover, \( x^* = \tilde{a} - \tilde{b} \), where \( \tilde{a} = (22.4983, 0.8118) \in E_1 + E_2 \) and \( \tilde{b} = (-2.9236, 10.5820) \in E_3 \) is the pair of closest points; see Figure 6.

**Example 7.4** We now consider the problem of computing the projection of the origin onto a Minkowski sum of two ellipsoids \( E_1(A_1, c_1) \) and \( E_2(A_2, c_2) \) in high dimensions. Let

\[
M = \left\{ 10^{(i-1)\text{cond}} \mid i = 1, \ldots, d \right\}
\]

where \( d \) is the space dimension and the number \( \text{cond} \) allows us to adjust the shapes (thin or fat) of the ellipsoids. For each pair \( (d, \text{cond}) \), we generate 1000 problems and implement both NESMINO and Gilbert algorithm, and compute the average CPU time in seconds. In each problem, let \( A_1 \) and \( A_2 \) are \( d \times d \) diagonal matrices such that the main diagonal entries of each of them is some permutation of \( M \). Since we want to guarantee that \( 0 \notin E_1 + E_2 \), we choose the two corresponding centers \( c_1, c_2 \in \mathbb{R}^{d \times 1} \) such that each of their entries is chosen randomly between \( m \) and \( 11m \), where \( m = \sqrt{\frac{10^{d\text{cond}}}{d}} \).

The result is reported in Table 1.

To compare the accuracy of both algorithms, for each \( i \)th problem among 1000 problems corresponding to a fix pair \( (d, \text{cond}) \), we also save the objective function value at final iteration of the two methods by \( f_{\text{NESMINO}}(i) \) and \( f_G(i) \), and count how many \( i \) such that \( |f_{\text{NESMINO}}(i) - f_G(i)| < 10^{-6} \). We see that almost 1000 problems in each pair \( (d, \text{cond}) \) satisfying this check. From Table 1, we can observe that the CPU time almost increases with \( d \) and \( \text{cond} \). Moreover, the smoothing algorithm depends more heavily on the shapes of ellipsoids than Gilbert’s algorithm. The Table also show that both algorithm may have good potential for solving large scale problems. This example also show that despite conservative theoretical bound on the rate of convergence, in the case of ellipsoids, Gilbert’s algorithm turns to be faster than smoothing algorithm.

### 8 Conclusions

Minimum norm problems have been studied from both theoretical and numerical point of view in this paper. Based on the minimum norm duality theorem, it is shown that
Table 1: Performance of NESMINO algorithm and Gilbert’s algorithm in solving minimum norm problems associated with ellipsoids.

| d  | cond | NESMINO   | Gilbert  |
|----|------|-----------|----------|
| 10 | 2    | 0.0021    | 0.0007   |
| 10 | 3    | 0.0006    | 0.0006   |
| 10 | 4    | 0.0006    | 0.0006   |
| 5  | 5    | 0.0006    | 0.0006   |
| 100| 2    | 0.0006    | 0.0006   |
| 100| 3    | 0.0006    | 0.0006   |
| 100| 4    | 0.0006    | 0.0006   |
| 100| 5    | 0.0006    | 0.0006   |
| 200| 2    | 0.0006    | 0.0006   |
| 200| 3    | 0.0006    | 0.0006   |
| 200| 4    | 0.0006    | 0.0006   |
| 200| 5    | 0.0006    | 0.0006   |
| 500| 2    | 0.0006    | 0.0006   |
| 500| 3    | 0.0006    | 0.0006   |
| 500| 4    | 0.0006    | 0.0006   |
| 500| 5    | 0.0006    | 0.0006   |

projections onto a Minkowski sum of sets can be represented as the sum of points on constituent sets so that, at these points, all of the sets share the same normal vector. By combining Nesterov’s smoothing technique and his fast gradient scheme, we have developed a numerical algorithm for solving the problems. The proposed algorithm is proved to have a better convergence rate than Gilbert’s algorithm in the worst case. Numerical examples also show that the algorithm works well for the problem in high dimensions. We also note that Gilbert’s algorithm is a Frank-Wolf type method; see [9, 15]. Although the convergence rate is known not to be very fast, due to its very cheap computational cost per iteration, its variants are still methods of choice in many applications.

References

[1] G. van den Bergen, *A fast and robust GJK implementation for collision detection of convex objects*, Tech. report, Department of Mathematics and Computing Science, Eindhoven University of Technology, 1999.

[2] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, CMS books in Mathematics. Canadian Mathematical Society, 2000.

[3] S. Cameron, *Enhancing GJK: Computing minimum and penetration distances between convex polyhedra*, In Proceedings of International Conference on Robotics and Automation, 3112-3117, 1997.

[4] Y. Chen and X. Ye, *Projection onto a simplex*, CoRR, abs/1208.4873

[5] L. Condat, *Fast projection onto the simplex and the ℓ₁ ball*, Math. Program., 158 (2016), 575-585.
[6] Y. H. Dai, *Fast algorithms for projection on an ellipsoid*, SIAM J. Optim., 16 (2006), 986-1006.

[7] A. Dax, *A new class of minimum norm duality theorems*, SIAM J. Optim., 19 (2009) 1947-1969.

[8] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, *Efficient projections onto the $\ell_1$-ball for learning in high dimensions*, in Proceedings of the 25th ACM International Conference on Machine learning, 2008, 272-279.

[9] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Res. Logis. Quart., 3 (1956), 95-110.

[10] E. G. Gilbert, *An iterative procedure for computing the minimum of a quadratic form on a convex set*, SIAM J. Contr., 4 (1966), 61-80.

[11] E. G. Gilbert, D. W. Johnson, S. S. Keerthi, *A fast procedure for computing the distance between complex objects in three-dimensional space*, IEEE Trans. Robot. Autom., 4 (1988), 193-203.

[12] E. G. Gilbert and C.-P. Foo, *Computing the distance between general convex objects in three-dimensional space*, IEEE Trans. Robot. Autom. 6 (1990), 53-61.

[13] Z. R. Gabidullina, *The problem of projecting the origin of euclidean space onto the convex polyhedron*, [http://arxiv.org/abs/1605.05351](http://arxiv.org/abs/1605.05351).

[14] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, I and II, Grundlehren Math. Wiss. 305 and 306. Springer-Verlag, Berlin, 1993.

[15] M. Jaggi, *Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization*, In ICML (1), pp. 427-435, 2013.

[16] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, *A fast iterative nearest point algorithm for support vector machine classifier design*, IEEE Trans. Neural Netw., 11 (2000), 124-136.

[17] A. A. Kurzhanskiy and P. Varaiya, *Ellipsoidal Toolbox*, Tech. Report EECS-2006-46, EECS, UC Berkeley, 2006.

[18] M. Kvasnica, P. Grieder, M. Baotic, and M. Morari, *Multi-Parametric Toolbox (MPT)*, Automatic Control Laboratory, ETHZ, Zurich, 2003.

[19] D. G. Luenberger, *Optimization by Vector Spaces Method*, John Wiley and Sons, Inc., New York, 1969.

[20] S. Martin, *Training support vector machines using Gilbert’s algorithm*, The 5th IEEE International Conference on Data Mining (ICDM), 2005, 306-313.
[21] B.F. Mitchell, V.F. Demyanov, and V.N. Malozemov, *Finding the point of a polyhedron closest to the origin*, SIAM J. Control Optim., 12 (1974), 19-26.

[22] B. S. Mordukhovich and N. M. Nam, *Limiting subgradients of minimal time functions in Banach spaces*, J. Global Optim., 46 (2010), 615-633.

[23] N. M. Nam, N. T. An, R. B. Rector, and J. Sun, *Nonsmooth algorithms and Nesterov smoothing technique for generalized Fermat-Torricelli problems*, SIAM J. Optim., 24 (2014), No. 4, 1815-1839.

[24] Yu. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program. 103 (2005), 127-152.

[25] Yu. Nesterov, *Introductory lectures on convex optimization. A basic course*, Appl. Optim. 87, Kluwer Academic Publishers, Boston, MA, 2004.

[26] Yu. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $O\left(\frac{1}{k^2}\right)$, Dokl. Akad. Nauk SSSR, 269 (1983), 543-547.

[27] L. Nirenberg, *Functional Analysis*, Academic Press, New York, 1961.

[28] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[29] H. Tuy, *Convex Analysis and Global Optimization. Nonconvex Optimization and Its Applications*, Kluwer Academic Publishers, 1998.

[30] P. Wolfe, *Finding the nearest point in a polytope*, Math. Program., 11 (1976), 128-149.