SIAM J. OPTIM.
Vol. 24, No. 4, pp. 1815–1839

NONSMOOTH ALGORITHMS AND NESTEROV'S SMOOTHING TECHNIQUE FOR GENERALIZED FERMAT–TORRICELLI PROBLEMS*

NGUYEN MAU NAM†, NGUYEN THAI AN‡, R. BLAKE RECTOR†, AND JIE SUN§

Abstract. We present algorithms for solving a number of new models of facility location which generalize the classical Fermat–Torricelli problem. Our first approach involves using Nesterov's smoothing technique and the minimization majorization principle to build smooth approximations that are convenient for applying smooth optimization schemes. Another approach uses subgradient-type algorithms to cope directly with the nondifferentiability of the cost functions. Convergence results of the algorithms are proved and numerical tests are presented to show the effectiveness of the proposed algorithms.

Key words. MM principle, Nesterov's smoothing technique, Nesterov's accelerated gradient method, generalized Fermat–Torricelli problem, subgradient-type algorithms

AMS subject classifications. 49J52, 49K40, 58C20

DOI. 10.1137/130945442

1. Introduction.

The *Fermat–Torricelli problem* was introduced in the 17th century by the French mathematician Pierre de Fermat and can be stated as follows: Given a finite collection of points in the plane, find a point that minimizes the sum of the distances to those points. This practical problem has been the inspiration for many new problems in the fields of computational geometry, logistics, and location science. Many generalized versions of the Fermat–Torricelli have been introduced and studied over the years; see [14, 15, 17, 19, 20, 21, 26] and the references therein. In particular, the generalized Fermat–Torricelli problems involving distances to sets were the topics of recent research; see [2, 4, 7, 19, 20].

In this paper, we focus mainly on developing effective numerical algorithms for generalized Fermat–Torricelli problems. Let $\mathbb{R}^n$ be the n-dimensional Euclidean space. Given a nonempty compact convex set $F \subset \mathbb{R}^n$ that contains the origin as an interior point, define the function

$$\sigma_F(u) := \sup\{\langle u, x\rangle \mid x \in F\}, \tag{1}$$

which reduces to the *dual norm* generated by a norm $\|\cdot\|_X$ when $F := \{x \in \mathbb{R}^n \mid \|x\|_X \le 1\}$.

The *generalized distance function* defined by the *dynamic* F and the *target set* Θ is given by

$$d_F(x;\Theta) := \inf\{\sigma_F(x-w) \mid w \in \Theta\}. \tag{2}$$

*Received by the editors November 15, 2013; accepted for publication (in revised form) July 25, 2014; published electronically October 28, 2014.
http://www.siam.org/journals/siopt/24-4/94544.html

†Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR 97207 (mau.nam.nguyen@pdx.edu, r.b.rector@pdx.edu). The research of the first author was partially supported by the USA National Science Foundation under grant DMS-1411817 and the Simons Foundation under grant 208785.

‡Thua Thien Hue College of Education, Hue City, Vietnam (thaian2784@gmail.com).
§Department of Mathematics and Statistics, Curtin University, Perth 6845, Australia (jie.sun@curtin.edu.au).

If F is the closed unit Euclidean ball of $\mathbb{R}^n$, the function (2) reduces to the *shortest distance function* or simply the *distance function*

$$d(x;\Theta) := \inf\{\|x-w\| \mid w \in \Theta\}. \tag{3}$$

Given a finite collection of nonempty closed convex sets Ω_i for $i = 1,\dots,m$, consider the following optimization problem:

$$\text{minimize } T(x),\ x \in \Omega, \tag{4}$$

where Ω is a convex constraint set, and the cost function T is defined by

$$T(x) := \sum_{i=1}^{m} d_F(x;\Omega_i).$$

In the general case of problem (4), the objective function T is not necessarily smooth. To solve problem (4) or, more generally, a nonsmooth optimization problem, a natural idea involves using smoothing techniques to approximate the original nonsmooth problem by a smooth one. Then, different smooth optimization schemes are applied to the smooth problem. One of the successful implementations of this idea was provided by Nesterov. In his seminal papers [25, 23], Nesterov introduced a fast first-order method for solving convex smooth optimization problems in which the cost functions have Lipschitz gradient. In contrast to the convergence rate of $O(1/k)$ when applying the classical gradient method to this class of problems, Nesterov's accelerated gradient method gives a convergence rate of $O(1/k^2)$. In Nesterov's nonsmooth optimization scheme, an original nonsmooth function of a particular form is approximated by a smooth convex function with Lipschitz gradient. Then the accelerated gradient method can be applied to solve the smooth approximation.

Another approach uses subgradient-type algorithms to cope directly with the nondifferentiability. In fact, subgradient-type algorithms allow us to solve the problem in very broad settings that involve distance functions generated by different norms and also generalized distance functions generated by different sets F. However, the classical subgradient method is known to be slow in general. Thus, it is not a good option when the number of target sets is large in high dimensions. We apply the *stochastic subgradient method* to deal with this situation. It has been shown that the stochastic subgradient method is an effective tool for solving many practical problems; see [1, 28] and the references therein. This simple method also shows its effectiveness for solving the generalized Fermat–Torricelli problem.

The remainder of this paper is organized as follows. In section 2 we give an introduction to Nesterov's smoothing technique, Nesterov's accelerated gradient method, and the minimization majorization (MM) principle to solve nonsmooth optimization problems. These tools will be used in sections 3 and 4 to develop numerical algorithms for solving generalized Fermat–Torricelli problems with points and sets. Subgradient-type algorithms for solving these problems are also presented in section 4. Section 5 contains numerical examples to illustrate the algorithms.

Throughout the paper, $\langle\cdot,\cdot\rangle$ denotes the usual inner product in $\mathbb{R}^n$, and the corresponding Euclidean norm is denoted by $\|\cdot\|$; F is assumed to be a nonempty compact convex set in $\mathbb{R}^n$ that contains 0 as an interior point; $\operatorname{bd} F$ denotes the topological boundary of F. We also use basic concepts and results of convex optimization, which can be found, e.g., in [24, 27].

2. Nesterov's smoothing technique, accelerated gradient method, and MM principle. In this section we study and provide more details on Nesterov's smoothing technique and accelerated gradient method introduced in [23]. We also present a general form of the MM principle well known in computational statistics.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. Consider the constrained optimization problem

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

where f is not necessarily differentiable and Ω is a nonempty closed convex subset of $\mathbb{R}^n$.

The class of functions under consideration is given by

$$f(x) := \max\{\langle Ax, u\rangle - \phi(u) \mid u \in Q\},\ x \in \mathbb{R}^n,$$

where A is an $m \times n$ matrix, Q is a nonempty compact convex subset of $\mathbb{R}^m$, and ϕ is a continuous convex function on Q.

Let d be a continuous *strongly convex function* on Q with parameter $\sigma > 0$. The function d is called a *prox-function*. Since d is strongly convex on Q, it has a unique optimal solution on this set. Denote

$$\bar{u} := \arg\min\{d(u) \mid u \in Q\}.$$

Without loss of generality, we assume that $d(\bar{u}) = 0$. From the strong convexity of d, we also have

$$d(u) \geq \frac{\sigma}{2}\|u - \bar{u}\|^2 \quad \text{for all } u \in Q.$$

Throughout the paper we will work mainly with the case where $d(u) = \frac{1}{2}\|u - \bar{u}\|^2$.

Let μ be a positive number called a *smooth parameter*. Define

$$f_\mu(x) := \max\{\langle Ax, u\rangle - \phi(u) - \mu d(u) \mid u \in Q\}. \tag{5}$$

The function f_μ will be the smooth approximation of f.

For an $m \times n$ matrix $A = (a_{ij})$, define

$$\|A\| := \max\{\|Ax\| \mid \|x\| \leq 1\}. \tag{6}$$

The definition gives us

$$\|Ax\| \leq \|A\|\, \|x\| \quad \text{for all } x \in \mathbb{R}^n.$$

We also recall the definition of the *Euclidean projection* from point $x \in \mathbb{R}^n$ to a nonempty closed convex subset Ω of $\mathbb{R}^n$:

$$\pi(x; \Omega) := \{w \in \Omega \mid d(x; \Omega) = \|x - w\|\}.$$

Let us present below a simplified version of [23, Theorem 1] that involves the usual inner product of $\mathbb{R}^n$. We provide a new detailed proof for the convenience of the reader.

Theorem 2.1. *Consider the function f given by*

$$f(x) := \max\{\langle Ax, u\rangle - \langle b, u\rangle \mid u \in Q\},\ x \in \mathbb{R}^n,$$

where A is an $m \times n$ matrix and Q is a compact subset of $\mathbb{R}^m$. Let $d(u) = \frac{1}{2}\|u - \bar{u}\|^2$ with $\bar{u} \in Q$.

Then the function f_μ in (5) has the explicit representation

$$f_\mu(x) = \frac{\|Ax - b\|^2}{2\mu} + \langle Ax - b, \bar{u}\rangle - \frac{\mu}{2}\left[d(\bar{u} + \frac{Ax - b}{\mu}; Q)\right]^2$$

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

$$\nabla f_\mu(x) = A^\top u_\mu(x),$$

where u_μ can be expressed in terms of the Euclidean projection

$$u_\mu(x) = \pi\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right).$$

The gradient ∇f_μ is a Lipschitz function with constant

$$\ell_\mu = \frac{1}{\mu}\|A\|^2.$$

Moreover,

$$f_\mu(x) \le f(x) \le f_\mu(x) + \frac{\mu}{2}[D(\bar{u}; Q)]^2 \quad \text{for all } x \in \mathbb{R}^n, \tag{7}$$

where $D(\bar{u}; Q)$ is the farthest distance from $\bar{u}$ to Q defined by

$$D(\bar{u}; Q) := \sup\{\|\bar{u} - u\| \mid u \in Q\}.$$

Proof. We have

$$\begin{aligned}
f_\mu(x) &= \sup\left\{\langle Ax - b, u\rangle - \frac{\mu}{2}\|u - \bar{u}\|^2 \mid u \in Q\right\}\\
&= \sup\left\{-\frac{\mu}{2}\left(\|u - \bar{u}\|^2 - \frac{2}{\mu}\langle Ax - b, u\rangle\right) \mid u \in Q\right\}\\
&= -\frac{\mu}{2}\inf\left\{\left\|u - \bar{u} - \frac{Ax - b}{\mu}\right\|^2 - \frac{\|Ax - b\|^2}{\mu^2} - \frac{2}{\mu}\langle Ax - b, \bar{u}\rangle \mid u \in Q\right\}\\
&= \frac{\|Ax - b\|^2}{2\mu} + \langle Ax - b, \bar{u}\rangle - \frac{\mu}{2}\inf\left\{\left\|u - \bar{u} - \frac{Ax - b}{\mu}\right\|^2 \mid u \in Q\right\}\\
&= \frac{\|Ax - b\|^2}{2\mu} + \langle Ax - b, \bar{u}\rangle - \frac{\mu}{2}\left[d\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right)\right]^2.
\end{aligned}$$

Since the function $\psi(u) := [d(u; Q)]^2$ is continuously differentiable with $\nabla\psi(u) = 2[u - \pi(u; Q)]$ for all $u \in \mathbb{R}^m$ (see, e.g., [11, p. 186]), it follows from the chain rule that

$$\begin{aligned}
\nabla f_\mu(x) &= \frac{1}{\mu}A^\top(Ax - b) + A^\top\bar{u} - \frac{\mu}{2}\left[\frac{2}{\mu}A^\top\left(\bar{u} + \frac{Ax - b}{\mu} - \pi\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right)\right)\right]\\
&= A^\top\pi\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right).
\end{aligned}$$

From the property of the projection mapping (see [11, Proposition 3.1.3, p. 118]) and the Cauchy–Schwarz inequality, for any $x, y \in \mathbb{R}^n$ we have

$$\begin{aligned}
\|\nabla f_\mu(x) - \nabla f_\mu(y)\|^2 &= \left\|A^\top \pi\left(\bar{u} + \frac{Ax-b}{\mu}; Q\right) - A^\top \pi\left(\bar{u} + \frac{Ay-b}{\mu}; Q\right)\right\|^2 \\
&\le \|A\|^2 \left\|\pi\left(\bar{u} + \frac{Ax-b}{\mu}; Q\right) - \pi\left(\bar{u} + \frac{Ay-b}{\mu}; Q\right)\right\|^2 \\
&\le \|A\|^2 \left\langle \frac{Ax-Ay}{\mu}, \pi\left(\bar{u} + \frac{Ax-b}{\mu}; Q\right) - \pi\left(\bar{u} + \frac{Ay-b}{\mu}; Q\right)\right\rangle \\
&= \frac{\|A\|^2}{\mu} \left\langle x-y, A^\top \pi\left(\bar{u} + \frac{Ax-b}{\mu}; Q\right) - A^\top \pi\left(\bar{u} + \frac{Ay-b}{\mu}; Q\right)\right\rangle \\
&= \frac{\|A\|^2}{\mu} \langle x-y, \nabla f_\mu(x) - \nabla f_\mu(y)\rangle \\
&\le \frac{\|A\|^2}{\mu} \|x-y\| \|\nabla f_\mu(x) - \nabla f_\mu(y)\|.
\end{aligned}$$

This implies that

$$\|\nabla f_\mu(x) - \nabla f_\mu(y)\| \le \frac{\|A\|^2}{\mu}\|x-y\|.$$

The lower and upper bounds in (7) follow from

$$\langle Ax-b, u\rangle - \frac{\mu}{2}\|u-\bar{u}\|^2 \le \langle Ax-b, u\rangle \le \left(\langle Ax-b, u\rangle - \frac{\mu}{2}\|u-\bar{u}\|^2\right) + \frac{\mu}{2}\sup\{\|q-\bar{u}\|^2 \mid q \in Q\}$$

for all $x \in \mathbb{R}^n$ and $u \in Q$. $\square$

Example 2.2. Let $\|\cdot\|_{X_1}$ and $\|\cdot\|_{X_2}$ be two norms in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and let $\|\cdot\|_{X_1^*}$ and $\|\cdot\|_{X_2^*}$ be the corresponding dual norms, i.e.,

$$\|x\|_{X_i^*} := \sup\{\langle x, u\rangle \mid \|u\|_{X_i} \le 1\},\ i = 1, 2.$$

Denote $\mathbb{B}_{X_1} := \{u \in \mathbb{R}^m \mid \|u\|_{X_1} \le 1\}$ and $\mathbb{B}_{X_2} := \{u \in \mathbb{R}^n \mid \|u\|_{X_2} \le 1\}$. Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$g(x) := \|Ax-b\|_{X_1^*} + \lambda\|x\|_{X_2^*},$$

where A is an $m\times n$ matrix, $b \in \mathbb{R}^m$, and $\lambda > 0$. Using the prox-function $d(u) = \frac{1}{2}\|u\|^2$, one finds a smooth approximation of f below:

$$g_\mu(x) = \frac{\|Ax-b\|^2}{2\mu} - \frac{\mu}{2}\left[d\left(\frac{Ax-b}{\mu}; \mathbb{B}_{X_1}\right)\right]^2 + \lambda\left(\frac{\|x\|^2}{2\mu} - \frac{\mu}{2}\left[d\left(\frac{x}{\mu}; \mathbb{B}_{X_2}\right)\right]^2\right).$$

The gradient of f_μ is

$$\nabla g_\mu(x) = A^\top \pi\left(\frac{Ax-b}{\mu}; \mathbb{B}_{X_1}\right) + \lambda\pi\left(\frac{x}{\mu}; \mathbb{B}_{X_2}\right),$$

and its Lipschitz constant is

$$L_\mu = \frac{\|A\|^2 + \lambda}{\mu}.$$

Moreover,

$$g_\mu(x) \le g(x) \le g_\mu(x) + \frac{\mu}{2}([D(0;\mathbb{B}_{X_1})]^2 + [D(0;\mathbb{B}_{X_2})]^2) \quad \text{for all } x \in \mathbb{R}^n.$$

For example, if $\|\cdot\|_{X_1}$ is the Euclidean norm, and $\|\cdot\|_{X_2}$ is the ℓ^∞-norm on $\mathbb{R}^n$, then

$$\nabla g_\mu(x) = A^\top \frac{Ax-b}{\max\{\|Ax-b\|,\mu\}} + \lambda \text{median}\left(\frac{x}{\mu}, e, -e\right),$$

where $e = [1,\dots,1]^\top \in \mathbb{R}^n$.

Let us provide another example of *support vector machine problems*. Our approach simplifies and improves the results in [32].

Example 2.3. Let $S := \{(X_i, y_i)\}_{i=1}^m$ be a training set, where $X_i \in \mathbb{R}^p$ is the ith row of a matrix X and $y_i \in \{-1,1\}$. The corresponding linear support vector machine problem can be reduced to solving the following problem:

$$\text{minimize } g(w) := \frac{1}{2}\|w\|^2 + \lambda \sum_{i=1}^m \ell_i(w), w \in \mathbb{R}^p,$$

where $\ell_i(w) = \max\{0, 1 - y_i X_i w\}$, $\lambda > 0$.

Let $Q := \{u \in \mathbb{R}^m \mid 0 \le u_i \le 1\}$, and define

$$f(w) := \sum_{i=1}^m \ell_i(w) = \max_{u\in Q}\langle e - YXw, u\rangle,$$

where $e = [1,\dots,1]^\top$ and $Y = \text{diag}(y)$ with $y = [y_1,\dots,y_m]^\top$.

Using the prox-function $d(u) = \frac{1}{2}\|u\|^2$, one has

$$f_\mu(w) = \max_{u\in Q}[\langle e - YXw, u\rangle - \mu d(u)].$$

Then

$$u_\mu(w) = \pi\left(\frac{e - YXw}{\mu}; Q\right) = \left\{u \in \mathbb{R}^m \mid u_i = \text{median}\left\{\frac{1 - y_i X_i w}{\mu}, 0, 1\right\}\right\}.$$

The gradient of f_μ is given by

$$\nabla f_\mu(w) = -(YX)^\top u_\mu(w),$$

and its Lipschitz constant is $\ell_\mu = \frac{\|YX\|^2}{\mu}$, where the matrix norm is defined in (6).

Moreover,

$$f_\mu(w) \le f(w) \le f_\mu(w) + \frac{m\mu}{2} \quad \text{for all } w \in \mathbb{R}^p.$$

Then we use the following smooth approximation of the original objective function g:

$$g_\mu(w) := \frac{1}{2}\|w\|^2 + \lambda f_\mu(w), \quad w \in \mathbb{R}^p.$$

Obviously,

$$\nabla g_\mu(w) = w + \lambda \nabla f_\mu(w),$$

and a Lipschitz constant is

$$L_\mu = 1 + \lambda \frac{\|YX\|^2}{\mu}.$$

The smooth approximations obtained above are convenient for applying Nesterov's accelerated gradient method presented in what follows. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function with Lipschitz gradient. That is, there exists $\ell > 0$ such that

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

Let Ω be a nonempty closed convex set. In his paper [23], Nesterov considered the optimization problem

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

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

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

Let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a strongly convex function with parameter $\sigma > 0$, and let $x_0 \in \mathbb{R}^n$ such that

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

Further, assume that $\rho(x_0) = 0$. Then Nesterov's accelerated gradient algorithm is outlined as follows.

Algorithm 1.

INPUT: f, ℓ.
INITIALIZE: Choose $x_0 \in \Omega$.
Set $k = 0$
Repeat the following
  Find $y_k := T_\Omega(x_k)$.
  Find $z_k := \arg\min\big\{\frac{\ell}{\sigma}\rho(x) + \sum_{i=0}^{k} \frac{i+1}{2}[f(x_i) + \langle \nabla f(x_i), x - x_i\rangle] \,\big|\, x \in \Omega\big\}$.
  Set $x_{k+1} := \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$.
  Set $k := k + 1$.
until a stopping criterion is satisfied.
OUTPUT: y_k.

For simplicity, we choose $\rho(x) = \frac{\sigma}{2}\|x - x_0\|^2$, where $x_0 \in \Omega$ and $\sigma = 1$. Following the proof of Theorem 2.1, it is not hard to see that

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

Moreover,

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

We continue with another important tool of convex optimization and computational statistics called the *MM principle*; see [8, 12, 16] and the references therein.

Here we provide a more general version. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and let Ω be a nonempty closed convex subset of $\mathbb{R}^n$. Consider the optimization problem

$$\text{minimize } f(x) \text{ subject to } x \in \Omega. \tag{8}$$

Let $\mathcal{M} : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, and let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ be a set-valued mapping with nonempty values such that the following properties hold for all $x, y \in \mathbb{R}^n$:

$$f(x) \le \mathcal{M}(x, z) \quad \text{for all } z \in F(y), \text{ and } f(x) = \mathcal{M}(x, z) \quad \text{for all } z \in F(x).$$

Given $x_0 \in \Omega$, the MM algorithm to solve (8) is given by

$$\text{Choose } z_k \in F(x_k) \text{ and find } x_{k+1} \in \arg\min\{\mathcal{M}(x, z_k) \mid x \in \Omega\}.$$

Then

$$f(x_{k+1}) \le \mathcal{M}(x_{k+1}, z_k) \le \mathcal{M}(x_k, z_k) = f(x_k).$$

Finding an appropriate majorization is an important step in this algorithm. It has been shown in [7] that the MM principle provides an effective tool for solving the generalized Fermat–Torricelli problem. In what follows, we apply the MM principle in combination with Nesterov's smoothing technique and accelerated gradient method to solve generalized Fermat–Torricelli problems in many different settings.

3. Generalized Fermat–Torricelli problems involving points.

Let Ω be a nonempty closed convex subset of $\mathbb{R}^n$, and let $a_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$. In this section, we consider the following generalized version of the Fermat–Torricelli problem:

$$\text{minimize } H(x) := \sum_{i=1}^{m} \sigma_F(x - a_i) \text{ subject to } x \in \Omega. \tag{9}$$

Let us start with some properties of the function σ_F used in problem (9). The following proposition can be proved easily.

PROPOSITION 3.1. *For the function σ_F defined in* (1), *the following properties hold for all $u, v \in \mathbb{R}^n$ and $\lambda \ge 0$:*

(i) $|\sigma_F(u) - \sigma_F(v)| \le \|F\|\|u - v\|$, *where* $\|F\| := \sup\{\|f\| \mid f \in F\}$.
(ii) $\sigma_F(u + v) \le \sigma_F(u) + \sigma_F(v)$.
(iii) $\sigma_F(\lambda u) = \lambda\sigma_F(u)$, *and* $\sigma_F(u) = 0$ *if and only if* $u = 0$.
(iv) σ_F *is a norm if we assume additionally that F is symmetric, i.e.,* $F = -F$.
(v) $\gamma\|u\| \le \sigma_F(u)$, *where* $\gamma := \sup\{r > 0 \mid \mathbb{B}(0; r) \subset F\}$.

Let Θ be a nonempty closed convex subset of $\mathbb{R}^n$, and let $\bar{x} \in \Theta$. The *normal cone* in the sense of convex analysis to Θ at $\bar{x}$ is defined by

$$N(\bar{x}; \Theta) := \{v \in \mathbb{R}^n \mid \langle v, x - \bar{x}\rangle \le 0 \text{ for all } x \in \Theta\}.$$

It follows from the definition that the normal cone mapping $N(\cdot; \Theta)$ has closed graph in the sense that for any sequence $x_k \to \bar{x}$ and $v_k \to \bar{v}$ where $v_k \in N(x_k; \Theta)$, one has that $\bar{v} \in N(\bar{x}; \Theta)$.

Given an element $v \in \mathbb{R}^n$, we also define $\operatorname{cone}\{v\} := \{\lambda v \mid \lambda \ge 0\}$.

In what follows, we study the existence and uniqueness of the optimal solution of problem (9). The following definition and the proposition afterward are important for this purpose.

DEFINITION 3.2. *We say that F is* normally smooth *if for every $x \in \operatorname{bd} F$ there exists $a_x \in \mathbb{R}^n$ such that $N(x; F) = \operatorname{cone}\{a_x\}$.*

Given a positive definite matrix A, let

$$\|x\|_X := \sqrt{x^\top Ax}.$$

It is not hard to see that the set $F := \{x \in \mathbb{R}^n \mid \|x\|_X \le 1\}$ is normally smooth. Indeed, $N(x;F) = \operatorname{cone}\{Ax\}$ if $\|x\|_X = 1$; see [18, Proposition 2.48].

Define the set

$$\mathbb{B}^*_F := \{u \in \mathbb{R}^n \mid \sigma_F(u) \le 1\},$$

and recall that a convex subset Θ of $\mathbb{R}^n$ is said to be *strictly convex* if $tu + (1-t)v \in \operatorname{int}\Theta$ whenever $u, v \in \Theta$, $u \ne v$, and $t \in (0,1)$.

PROPOSITION 3.3. *We have that F is normally smooth if and only if $\mathbb{B}^*_F$ is strictly convex.*

Proof. Suppose that F is normally smooth. Fix any $u, v \in \mathbb{B}^*_F$ with $u \ne v$ and $t \in (0,1)$. Let us show that $tu+(1-t)v \in \operatorname{int}\mathbb{B}^*_F$, or equivalently, $\sigma_F(tu+(1-t)v) < 1$. We need only consider the case where $\sigma_F(u) = \sigma_F(v) = 1$. Fix $\bar{x}, \bar{y} \in F$ such that

$$\langle u, \bar{x}\rangle = \sigma_F(u) = 1 \text{ and } \langle v, \bar{y}\rangle = \sigma_F(v) = 1,$$

and fix $e \in F$ such that

$$\langle tu + (1-t)v, e\rangle = \sigma_F(tu + (1-t)v).$$

It is obvious that $\sigma_F(tu + (1-t)v) \le 1$. By contradiction, suppose that $\sigma_F(tu + (1-t)v) = 1$. Then

$$1 = \langle tu + (1-t)v, e\rangle = t\langle u, e\rangle + (1-t)\langle v, e\rangle \le t\langle u, \bar{x}\rangle + (1-t)\langle v, \bar{y}\rangle = 1.$$

This implies that $\langle u, e\rangle = \langle u, \bar{x}\rangle = 1 = \sigma_F(u)$ and $\langle v, e\rangle = \langle v, \bar{y}\rangle = 1 = \sigma_F(v)$. Then

$$\langle u, x\rangle \le \langle u, e\rangle \quad \text{for all } x \in F,$$

which implies that $u \in N(e;F)$. Similarly, $v \in N(e;F)$. Since F is normally smooth, $u = \lambda v$, where $\lambda > 0$. Thus,

$$1 = \langle u, e\rangle = \langle \lambda v, e\rangle = \lambda\langle v, e\rangle = \lambda.$$

Hence $\lambda = 1$ and $u = v$, a contradiction.

Now suppose that $\mathbb{B}^*_F$ is strictly convex. Fix $\bar{x} \in \operatorname{bd} F$, and fix any $u, v \in N(\bar{x};F)$ with $u, v \ne 0$. Let $\alpha := \sigma_F(u)$ and $\beta := \sigma_F(v)$. Then

$$\langle u, x\rangle \le \langle u, \bar{x}\rangle \quad \text{for all } x \in F$$

and

$$\langle v, x\rangle \le \langle v, \bar{x}\rangle \quad \text{for all } x \in F.$$

It follows that $\langle u, \bar{x}\rangle = \alpha$ and $\langle v, \bar{x}\rangle = \beta$. Moreover,

$$\sigma_F(u + v) \ge \langle u, \bar{x}\rangle + \langle v, \bar{x}\rangle = \alpha + \beta = \sigma_F(u) + \sigma_F(v),$$

and hence $\sigma_F(u + v) = \sigma_F(u) + \sigma_F(v)$. We have $u/\alpha, v/\beta \in \mathbb{B}^*_F$ and

$$\sigma_F\left(\frac{u}{\alpha}\frac{\alpha}{\alpha+\beta} + \frac{v}{\beta}\frac{\beta}{\alpha+\beta}\right) = 1.$$

Since $\mathbb{B}_F^*$ is strictly convex, one has $\frac{u}{\alpha} = \frac{v}{\beta}$, and hence $u = \lambda v$, where $\lambda := \alpha/\beta > 0$. The proof is now complete. □

Remark 3.4. Suppose that F is normally smooth. It follows from the proof of Proposition 3.3 that for $u, v \in \mathbb{R}^n$ with $u, v \neq 0$, one has that $\sigma_F(u + v) = \sigma_F(u) + \sigma_F(v)$ if and only if $u = \lambda v$ for some $\lambda > 0$.

The proposition below gives sufficient conditions that guarantee the uniqueness of an optimal solution of (9).

PROPOSITION 3.5. *Suppose that F is normally smooth. If for any $x, y \in \Omega$ with $x \neq y$ the line connecting x and y, $\mathcal{L}(x, y)$, does not contain at least one of the points a_i for $i = 1, \ldots, m$, then problem* (9) *has a unique optimal solution.*

Proof. It is not hard to see that for any $\alpha \in \mathbb{R}$, the set

$$\mathcal{L}_\alpha := \{x \in \Omega \mid H(x) \leq \alpha\}$$

is compact, and so (9) has an optimal solution since H is continuous. Let us show that the assumptions made guarantee that H is strictly convex on Ω, and hence (9) has a unique optimal solution.

By contradiction, suppose that there exist $\bar{x}, \bar{y} \in \Omega$ with $\bar{x} \neq \bar{y}$ and $t \in (0, 1)$ such that

$$H(t\bar{x} + (1 - t)\bar{y}) = tH(\bar{x}) + (1 - t)H(\bar{y}).$$

Then

$$\begin{aligned}\sigma_F(t(\bar{x} - a_i) + (1 - t)(\bar{y} - a_i)) &= t\sigma_F(\bar{x} - a_i) + (1 - t)\sigma_F(\bar{y} - a_i) \\ &= \sigma_F(t(\bar{x} - a_i)) + \sigma_F((1 - t)(\bar{y} - a_i)) \quad \text{for all } i = 1, \ldots, m.\end{aligned}$$

If $\bar{x} = a_i$ or $\bar{y} = a_i$, then obviously $a_i \in \mathcal{L}(\bar{x}, \bar{y})$. Otherwise, by Remark 3.4, there exists $\lambda_i > 0$ such that

$$t(\bar{x} - a_i) = \lambda_i(1 - t)(\bar{y} - a_i).$$

This also implies that $a_i \in \mathcal{L}(\bar{x}, \bar{y})$. We have seen that $a_i \in \mathcal{L}(\bar{x}, \bar{y})$ for all $i = 1, \ldots, m$. This contradiction shows that (9) has a unique optimal solution. □

Let us consider the smooth approximation function given by

$$H_\mu(x) := \sum_{i=1}^{m} \left(\frac{\|x - a_i\|^2}{2\mu} + \langle x - a_i, \bar{u} \rangle - \frac{\mu}{2} \left[d\left(\bar{u} + \frac{x - a_i}{\mu}; F \right) \right]^2 \right), \tag{10}$$

where $\bar{u} \in F$.

PROPOSITION 3.6. *The function H_μ defined by* (10) *is continuously differentiable on $\mathbb{R}^n$ with its gradient given by*

$$\nabla H_\mu(x) = \sum_{i=1}^{m} \pi\left(\bar{u} + \frac{x - a_i}{\mu}; F \right).$$

The gradient ∇H_μ is a Lipschitz function with constant

$$\mathcal{L}_\mu = \frac{m}{\mu}.$$

Moreover, one has the following estimate:

$$H_\mu(x) \leq H(x) \leq H_\mu(x) + m\frac{\mu}{2}[D(\bar{u}; F)]^2 \quad \textit{for all } x \in \mathbb{R}^n.$$

Proof. Given $b \in \mathbb{R}^n$, define the function on $\mathbb{R}^n$ given by

$$f(x) := \sigma_F(x-b) = \max\{\langle x-b, u\rangle \mid u \in F\}, \quad x \in \mathbb{R}^n.$$

Consider the prox-function

$$d(u) := \frac{1}{2}\|u-\bar{u}\|^2.$$

Applying Theorem 2.1, one has that the function f_μ is continuously differentiable on $\mathbb{R}^n$ with its gradient given by

$$\nabla f_\mu(x) = u_\mu(x) = \pi\left(\bar{u} + \frac{x-b}{\mu}; F\right).$$

Moreover, the gradient ∇f_μ is a Lipschitz function with constant

$$\ell_\mu = \frac{1}{\mu}.$$

The explicit formula for f_μ is

$$f_\mu(x) = \frac{\|x-b\|^2}{2\mu} + \langle x-b, \bar{u}\rangle - \frac{\mu}{2}\left[d\left(\bar{u} + \frac{x-b}{\mu}; F\right)\right]^2.$$

The conclusions then follow easily. ∎

We are now ready to write a pseudocode for solving the Fermat–Torricelli problem (9).

Algorithm 2.

INPUT: a_i for $i = 1, \ldots, m$, μ.
INITIALIZE: Choose $x_0 \in \Omega$ and set $\ell = \frac{m}{\mu}$.
Set $k = 0$
Repeat the following
  Compute $\nabla H_\mu(x_k) = \sum_{i=1}^m \pi(\bar{u} + \frac{x_k - a_i}{\mu}; F)$.
  Find $y_k := \pi(x_k - \frac{1}{\ell}\nabla H_\mu(x_k); \Omega)$.
  Find $z_k := \pi(x_0 - \frac{1}{\ell}\sum_{i=0}^k \frac{i+1}{2}\nabla H_\mu(x_i); \Omega)$.
  Set $x_{k+1} := \frac{2}{k+3}z_k + \frac{k+1}{k+3}y_k$.
until a stopping criterion is satisfied.

Remark 3.7. When implementing Nesterov's accelerated gradient method, in order to get a more effective algorithm, instead of using a fixed smoothing parameter μ, we often change μ during the iteration. The general optimization scheme is

INITIALIZE: $x_0 \in \Omega$, $\mu_0 > 0$, $\mu_* > 0$, $\sigma \in (0, 1)$.
Set $k = 0$.
Repeat the following
  Apply Nesterov's accelerated gradient method with $\mu = \mu_k$ and starting point x_k to obtain an approximate solution x_{k+1}.
  Update $\mu_{k+1} = \sigma\mu_k$.
until $\mu \leq \mu_*$.

Example 3.8. In the case where F is the closed unit Euclidean ball, $\sigma_F(x) = \|x\|$ is the Euclidean norm and

$$\pi(x; F) = \begin{cases} \left\{\dfrac{x}{\|x\|}\right\}, & \|x\| > 1, \\ \{x\}, & \|x\| \leq 1. \end{cases}$$

Consider the ℓ_1-norm on $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$, one has

$$\|x\|_1 = \max\{\langle x, u\rangle \mid \|u\|_\infty \le 1\}.$$

In this case,

$$F = \{x \in \mathbb{R}^n \mid |x_i| \le 1 \text{ for all } i = 1, \ldots, n\}.$$

The smooth approximation of the function $f(x) := \|x\|_1$ depends on the Euclidean projection to the set F, which can be found explicitly. In fact, for any $u \in \mathbb{R}^n$, one has

$$\pi(u; F) = \{v \in \mathbb{R}^n \mid v_i = \text{median } \{u_i, 1, -1\}\}.$$

Now we consider the ℓ_∞-norm in $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$, one has

$$\|x\|_\infty = \max\{\langle x, u\rangle \mid \|u\|_1 \le 1\}.$$

In this case,

$$F = \{x \in \mathbb{R}^n \mid \|x\|_1 \le 1\}.$$

It is straightforward to find the Euclidean projection of a point to F in two and three dimensions. In the case of high dimensions, there are available algorithms to find an approximation of the projection; see, e.g., [10].

4. Generalized Fermat–Torricelli problems involving sets. In this section, we study generalized Fermat–Torricelli problems that involve sets. Consider the following optimization problem:

$$\text{minimize } T(x) := \sum_{i=1}^{m} d_F(x; \Omega_i) \text{ subject to } x \in \Omega, \tag{11}$$

where Ω and Ω_i for $i = 1, \ldots, m$ are nonempty closed convex sets and at least one of them is bounded. This assumption guarantees that the problem has an optimal solution. The sets Ω_i for $i = 1, \ldots, m$ are called the *target sets*, and the set Ω is called the *constraint set*.

The *generalized projection* from a point $x \in \mathbb{R}^n$ to a set Θ is defined based on the generalized distance function (2) as follows:

$$\pi_F(x; \Theta) := \{w \in \Theta \mid \sigma_F(x - w) = d_F(x; \Theta)\}. \tag{12}$$

Note that this set is not necessarily a singleton in general.

Before investigating problem (11), we study some important properties of the generalized distance function and the generalized projection to be used in what follows.

PROPOSITION 4.1. *Given a nonempty closed convex set* Θ, *consider the generalized distance function* (2) *and the generalized projection* (12). *The following properties hold:*

(i) *For* $\bar{x} \in \mathbb{R}^n$, *the set* $\pi_F(\bar{x}; \Theta)$ *is nonempty.*

(ii) *For* $\bar{x} \in \mathbb{R}^n$, $d_F(\bar{x}; \Theta) = 0$ *if and only if* $\bar{x} \in \Theta$.

(iii) *If* $\bar{x} \notin \Theta$ *and* $\bar{w} \in \pi_F(\bar{x}; \Theta)$, *then* $\bar{w} \in \operatorname{bd}\Theta$.

(iv) *If* F *is normally smooth, then* $\pi_F(\bar{x}; \Theta)$ *is a singleton for every* $\bar{x} \in \mathbb{R}^n$ *and the projection mapping* $\pi_F(\cdot; \Theta)$ *is continuous.*

Proof. The proofs of (i) and (ii) are straightforward.

(iii) Suppose by contradiction that $\bar{w} \in \operatorname{int}\Theta$. Choose $t \in (0,1)$ sufficiently small such that

$$w_t := \bar{w} + t(\bar{x} - \bar{w}) \in \Theta.$$

Then

$$\sigma_F(\bar{x} - w_t) = \sigma_F((1-t)(\bar{x} - \bar{w})) = (1-t)\sigma_F(\bar{x} - \bar{w}) = (1-t)d_F(\bar{x};\Theta) < d_F(\bar{x};\Theta),$$

which is a contradiction.

(iv) If $\bar{x} \in \Theta$, then $\pi_F(\bar{x};\Theta) = \{\bar{x}\}$. Consider the case where $\bar{x} \notin \Theta$. Suppose by contradiction that there exist $\bar{w}_1, \bar{w}_2 \in \pi_F(\bar{x};\Theta)$ with $\bar{w}_1 \neq \bar{w}_2$. Then

$$\gamma := \sigma_F(\bar{x} - \bar{w}_1) = \sigma_F(\bar{x} - \bar{w}_2) > 0.$$

By the positive homogeneity of σ_F,

$$\frac{\bar{x} - \bar{w}_1}{\gamma} \in \mathbb{B}_F^* \text{ and } \frac{\bar{x} - \bar{w}_2}{\gamma} \in \mathbb{B}_F^*.$$

From Proposition 3.3, the set $\mathbb{B}_F^*$ is strictly convex, and hence

$$\frac{1}{2}\left(\frac{\bar{x} - \bar{w}_1}{\gamma} + \frac{\bar{x} - \bar{w}_2}{\gamma}\right) \in \operatorname{int}\mathbb{B}_F^*.$$

This implies that

$$\frac{\bar{x} - (\bar{w}_1 + \bar{w}_2)/2}{\gamma} \in \operatorname{int}\mathbb{B}_F^*.$$

It follows again by the homogeneity of σ_F that

$$\sigma_F(\bar{x} - (\bar{w}_1 + \bar{w}_2)/2) < \gamma = d_F(\bar{x};\Theta),$$

which is a contradiction. It is not hard to show that $\pi_F(\cdot;\Theta)$ is continuous using a sequential argument by contradiction. ∎

To continue, we recall some basic concepts and results of convex analysis. A systematic development of convex analysis can be found, for instance, in [11, 18, 27]. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. For $\bar{x} \in \mathbb{R}^n$, a *subgradient* of f at $\bar{x}$ is a vector $v \in \mathbb{R}^n$ that satisfies

$$\langle v, x - \bar{x}\rangle \leq f(x) - f(\bar{x}) \quad \text{for all } x \in \mathbb{R}^n.$$

The set of all subgradients of f at $\bar{x}$ is called the *subdifferential* of f at this point and is denoted by $\partial f(\bar{x})$.

For a finite number of convex functions $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, m$, one has

$$\partial\left(\sum_{i=1}^m f_i\right)(x) = \sum_{i=1}^m \partial f_i(x), \qquad x \in \mathbb{R}^n.$$

It is well known that a convex function $f : \mathbb{R}^n \to \mathbb{R}$ has an absolute minimum on a convex set Ω at $\bar{x} \in \Omega$ if and only if

$$0 \in \partial f(\bar{x}) + N(\bar{x};\Omega).$$

The definition below is important in what follows.

DEFINITION 4.2. *A convex set F is said to be normally round if $N(x;F) \neq N(y;F)$ whenever $x, y \in \operatorname{bd} F$, $x \neq y$.*

PROPOSITION 4.3. *Given a nonempty closed convex set Θ, consider the generalized distance function* (2). *Then the following properties hold:*

(i) $|d_F(x;\Theta) - d_F(y;\Theta)| \leq \|F\| \, \|x-y\|$ *for all* $x, y \in \mathbb{R}^n$.

(ii) *The function $d_F(\cdot;\Theta)$ is convex, and for any $\bar{x} \in \mathbb{R}^n$,*

$$\partial d_F(\bar{x};\Theta) = \partial \sigma_F(\bar{x} - \bar{w}) \cap N(\bar{w};\Theta),$$

where $\bar{w} \in \pi_F(\bar{x};\Theta)$ and this representation does not depend on the choice of $\bar{w}$.

(iii) *If F is normally smooth and round, then the function $\sigma_F(\cdot)$ is differentiable at any nonzero point, and the function $d_F(\cdot;\Theta)$ is continuously differentiable on the complement of Θ with*

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

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

Proof. (i) This conclusion follows from the subadditivity and the Lipschitz property of the function σ_F.

(ii) The function $d_F(\cdot;\Theta)$ can be expressed as the following infimal convolution:

$$d_F(x;\Theta) = \inf\{\sigma_F(x-w) + \delta(w;\Theta) \mid w \in \mathbb{R}^n\} = (g \oplus \sigma_F)(x),$$

where $g(x) := \delta(x;\Theta)$ is the *indicator function* associated with Θ, i.e., $\delta(x;\Theta) = 0$ if $x \in \Theta$, and $\delta(x;\Theta) = \infty$ otherwise. For any $\bar{w} \in \pi_F(\bar{x};\Theta)$, one has

$$\sigma_F(\bar{x} - \bar{w}) + g(\bar{w}) = \sigma_F(\bar{x} - \bar{w}) = d_F(\bar{x};\Theta).$$

By [18, Corollary 2.65],

$$\partial d_F(\bar{x};\Theta) = \partial \sigma_F(\bar{x} - \bar{w}) \cap \partial g(\bar{w}) = \partial \sigma_F(\bar{x} - \bar{w}) \cap N(\bar{w};\Theta).$$

(iii) Let us first prove the differentiability of $\sigma_F(\cdot)$ at $\bar{x} \neq 0$. From [18, Theorem 2.68], one has

$$\partial \sigma_F(\bar{x}) = S(\bar{x}),$$

where $S(\bar{x}) := \{p \in F \mid \langle \bar{x}, p \rangle = \sigma_F(\bar{x})\}$. We will show that $S(\bar{x})$ is a singleton. By contradiction, suppose that there exist $p_1, p_2 \in S(\bar{x})$ with $p_1 \neq p_2$. From the definition, one has

$$\bar{x} \in N(p_1;F) = \operatorname{cone}\{a_1\} \text{ and } \bar{x} \in N(p_2;F) = \operatorname{cone}\{a_2\}.$$

Then there exist $\lambda_1, \lambda_2 > 0$ such that $\bar{x} = \lambda_1 a_1 = \lambda_2 a_2$, and hence $N(p_1;F) = N(p_2;F)$, a contradiction to the normally smooth and round properties of F. Thus, $\partial \sigma_F(\bar{x}) = S(\bar{x})$ is a singleton, and hence σ_F is differentiable at $\bar{x}$ by [18, Theorem 3.3].

Observe that the set $\partial d_F(\bar{x};\Theta)$ is always nonempty. Since $\partial d_F(\bar{x};\Theta) = \partial \sigma_F(\bar{x} - \bar{w}) \cap N(\bar{w};\Theta) = \nabla \sigma_F(\bar{x} - \bar{w}) \cap N(\bar{w};\Theta)$, it is obvious that $\nabla \sigma_F(\bar{x} - \bar{w}) \in N(\bar{w};\Theta)$ and $\partial d_F(\bar{x};\Theta) = \{\nabla \sigma_F(\bar{x} - \bar{w})\}$. Then the differentiability of $d_F(\cdot;\Theta)$ at $\bar{x}$ follows from [18, Theorem 3.3]. Since Θ^c is an open set and $\partial d_F(x;\Theta)$ is a singleton for every $x \in \Theta^c$, the function $d_F(\cdot;\Theta)$ is continuously differentiable on this set; see the corollary of [9, Proposition 2.2.2]. ∎

As a corollary, we obtain the following well-known formula for subdifferential of the distance function (3).

COROLLARY 4.4. *For a nonempty closed convex set* Θ*, the following representation holds for the distance function* (3)*:*

$$\partial d(\bar{x};\Theta) = \begin{cases} N(\bar{x};\Theta) \cap \mathbb{B} & \text{if } \bar{x} \in \Theta, \\ \left\{ \dfrac{\bar{x} - \pi(\bar{x};\Theta)}{d(\bar{x};\Theta)} \right\} & \text{if } \bar{x} \notin \Theta. \end{cases}$$

The following proposition gives sufficient conditions that guarantee the uniqueness of an optimal solution of problem (11).

PROPOSITION 4.5. *Suppose that* F *is normally smooth and the target sets* Ω_i *for* $i = 1,\dots,m$ *are strictly convex with at least one of them being bounded. If for any* $x, y \in \Omega$ *with* $x \neq y$ *the line connecting* x *and* y*,* $\mathcal{L}(x,y)$*, does not intersect at least one of the target sets, then problem* (11) *has a unique optimal solution.*

Proof. It is not hard to prove that if one of the target sets is bounded, then each level set $\{x \in \Omega \mid T(x) \leq \alpha\}$ is bounded. Thus, (11) has an optimal solution. It suffices to show that T is strictly convex on Ω under the given assumptions. By contradiction, suppose that T is not strictly convex. Then there exist $\bar{x}, \bar{y} \in \Omega$ and $t \in (0,1)$ with $\bar{x} \neq \bar{y}$ and

$$T(t\bar{x} + (1-t)\bar{y}) = tT(\bar{x}) + (1-t)T(\bar{y}).$$

This implies that $d_F(t\bar{x} + (1-t)\bar{y};\Omega_i) = td_F(\bar{x};\Omega_i) + (1-t)d_F(\bar{y};\Omega_i)$ for all $i = 1,\dots,m$. Choose $i_0 \in \{1,\dots,m\}$ such that $\mathcal{L}(\bar{x},\bar{y}) \cap \Omega_{i_0} = \emptyset$. Let $\bar{w}_1 := \pi_F(\bar{x};\Omega_{i_0})$ and $\bar{w}_2 := \pi_F(\bar{y};\Omega_{i_0})$. Then

$$\begin{aligned} d_F(t\bar{x} + (1-t)\bar{y};\Omega_{i_0}) &= td_F(\bar{x};\Omega_{i_0}) + (1-t)d_F(\bar{y};\Omega_{i_0}) \\ &= t\sigma_F(\bar{x} - \bar{w}_1) + (1-t)\sigma_F(\bar{y} - \bar{w}_2) \\ &\geq \sigma_F((t\bar{x} + (1-t)\bar{y}) - (t\bar{w}_1 + (1-t)\bar{w}_2)). \end{aligned}$$

It follows that $t\bar{w}_1 + (1-t)\bar{w}_2 = \pi_F(t\bar{x} + (1-t)\bar{y};\Omega_{i_0}) \in \operatorname{bd}\Omega_{i_0}$. By the strict convexity of Ω_{i_0}, one has $\bar{w}_1 = \bar{w}_2 =: \bar{w}$, and hence

$$\sigma_F(t(\bar{x} - \bar{w}) + (1-t)(\bar{y} - \bar{w})) = \sigma_F(t(\bar{x} - \bar{w})) + \sigma_F((1-t)(\bar{y} - \bar{w})).$$

Following the proof of Proposition 3.5 implies that $\bar{w} \in \mathcal{L}(\bar{x},\bar{y})$, a contradiction. ∎

Let us now apply the MM principle to the generalized Fermat–Torricelli problem. We rely on the following properties, which hold for all $x, y \in \mathbb{R}^n$:

(i) $d_F(x;\Theta) = \sigma_F(x - w)$ for all $w \in \pi_F(x;\Theta)$.

(ii) $d_F(x;\Theta) \leq \sigma_F(x - w)$ for all $w \in \pi_F(y;\Theta)$.

Consider the set-valued mapping

$$F(x) := \Pi_{i=1}^m \pi_F(x;\Omega_i).$$

Then cost function $T(x)$ is majorized by

$$T(x) \leq \mathcal{M}(x,w) := \sum_{i=1}^m \sigma_F(x - w^i),\ w = (w^1,\dots,w^m) \in F(y).$$

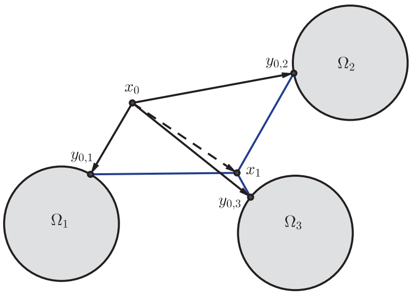

FIG. 1. *MM algorithm for a generalized Fermat–Torricelli problem.*

Moreover, $T(x) = \mathcal{M}(x, w)$ whenever $w \in F(x)$.

Thus, given $x_0 \in \Omega$, the MM iteration is given by

$$x_{k+1} \in \arg\min\{\mathcal{M}(x, w_k) \mid x \in \Omega\} \text{ with } w_k \in F(x_k).$$

This algorithm is illustrated in Figure 1 and can be written more explicitly as follows.

Algorithm 3.

INPUT: Ω and m target sets Ω_i, $i = 1, \ldots, m$.
INITIALIZE: $x_0 \in \Omega$.
Set $k = 0$.
Repeat the following
Find $y_{k,i} \in \pi_F(x_k; \Omega_i)$.
Solve the following problem with a stopping criterion $\min_{x \in \Omega} \sum_{i=1}^m \sigma_F(x - y_{k,i})$,
and denote its solution by x_{k+1}.
until a stopping criterion is satisfied.

Remark 4.6. Consider the Fermat–Torricelli problem

$$\text{minimize } \varphi(x) := \sum_{i=1}^{m} \|x - a_i\| \text{ subject to } x \in \Omega. \tag{13}$$

For $x \notin \{a_1, \ldots, a_m\}$,

$$\nabla\varphi(x) := \sum_{i=1}^{m} \frac{x - a_i}{\|x - a_i\|}.$$

Solving the equation $\nabla\varphi(x) = 0$ yields

$$x = \frac{\sum_{i=1}^m \dfrac{a_i}{\|x - a_i\|}}{\sum_{i=1}^m \dfrac{1}{\|x - a_i\|}}.$$

If $x \notin \{a_1, \ldots, a_m\}$, define

$$F(x) := \frac{\sum_{i=1}^m \dfrac{a_i}{\|x - a_i\|}}{\sum_{i=1}^m \dfrac{1}{\|x - a_i\|}}.$$

Otherwise, put $F(x) := x$. The Weiszfeld algorithm (see [13]) for solving problem (13) is stated as follows: Choose $x_0 \in \Omega$, and find $x_{k+1} := \pi(F(x_k); \Omega)$ for $k \geq 1$.

In the case where F is the closed unit Euclidean ball of $\mathbb{R}^n$ one has $\sigma_F(x) = \|x\|$. To solve the problem

$$\min_{x\in\Omega} \sum_{i=1}^{m} \|x - y_{k,i}\|$$

in the MM algorithm above, we can also use the Weiszfeld algorithm or its improvements.

PROPOSITION 4.7. *Consider the generalized Fermat–Torricelli problem* (11) *in which F is normally smooth and round. Let $\{x_k\}$ be the sequence in the MM algorithm defined by*

$$x_{k+1} \in \arg\min \left\{ \sum_{i=1}^{m} \sigma_F(x - \pi_F(x_k; \Omega_i)) \mid x \in \Omega \right\}.$$

Suppose that $\{x_k\}$ converges to $\bar{x}$ that does not belong to Ω_i for $i = 1, \ldots, m$. Then $\bar{x}$ is an optimal solution of problem (11).

Proof. Since the sequence $\{x_k\}$ converges to $\bar{x}$ that does not belong to Ω_i for $i = 1, \ldots, m$, we can assume that $x_k \notin \Omega_i$ for $i = 1, \ldots, m$ and for every k. From the definition of the sequence $\{x_k\}$, one has

$$0 \in \sum_{i=1}^{m} \nabla\sigma_F(x_{k+1} - \pi(x_k; \Omega_i)) + N(x_{k+1}; \Omega).$$

Using the continuity of $\nabla\sigma_F(\cdot)$ and the projection mapping $\pi_F(\cdot)$ to nonempty closed convex sets as well as the closedness of the normal cone mapping, one has

$$0 \in \sum_{i=1}^{m} \nabla\sigma_F(\bar{x} - \pi(\bar{x}; \Omega_i)) + N(\bar{x}; \Omega).$$

Thus,

$$0 \in \sum_{i=1}^{m} \partial d_F(\bar{x}; \Omega_i) + N(\bar{x}; \Omega) = \partial T(\bar{x}) + N(\bar{x}; \Omega).$$

It follows that $\bar{x}$ is also an optimal solution of problem (11). $\square$

It is of course important to find sufficient conditions that guarantee the convergence of the sequence $\{x_k\}$. This can be done using [8, Propositions 1 and 2]; see also [7]. We justify the use of this approach in the following lemma and apply it in the proposition that follows. For simplicity, we assume that the constraint set Ω does not intersect any of the target sets Ω_i for $i = 1, \ldots, m$.

LEMMA 4.8. *Consider the generalized Fermat–Torricelli problem* (11) *in which at least one of the target sets Ω_i for $i = 1, \ldots, m$ is bounded and F is normally smooth and round. Suppose that the constraint set Ω does not intersect any of the target sets Ω_i for $i = 1, \ldots, m$, and any $x, y \in \Omega$ with $x \neq y$ the line connecting x and y, $\mathcal{L}(x, y)$, does not intersect at least one of the target sets. For any $x \in \Omega$, consider the mapping $\psi : \Omega \to \Omega$ defined by*

$$\psi(x) := \arg\min \left\{ \sum_{i=1}^{m} \sigma_F(y - \pi_F(x; \Omega_i)) \mid y \in \Omega \right\}.$$

Then ψ is continuous at any point $\bar{x} \in \Omega$, and $T(\psi(x)) < T(x)$ whenever $x \neq \psi(x)$.

Proof. Fix any $x \in \Omega$. By Proposition 3.5 and from the assumptions made, the function

$$g(y) := \sum_{i=1}^{m} \sigma_F(y - \pi_F(x; \Omega_i))$$

is strictly convex on Ω, so $\psi(x)$ is the unique solution of the Fermat–Torricelli problem generated by $\pi_F(x; \Omega_i)$ for $i = 1, \ldots, m$. Thus, ψ is well defined. Fix any sequence $\{x_k\}$ that converges to $\bar{x}$. Then $y_k := \psi(x_k)$ satisfies

$$0 \in \sum_{i=1}^{m} \nabla \sigma_F(y_k - \pi_F(x_k; \Omega_i)) + N(y_k; \Omega).$$

Since at least one of the sets Ω_i for $i = 1, \ldots, m$ is bounded, we can show that the sequence $\{y_k\}$ is bounded. Indeed, suppose that Ω_1 is bounded and $\{y_k\}$ is unbounded. Then there exists a subsequence $\{y_{k_p}\}$ such that $\|y_{k_p}\| \to \infty$ as $p \to \infty$. For sufficiently large p and a fixed $y \in \Omega$, since $y_{k_p} = \arg\min\{\sum_{i=1}^{m} \sigma_F(y - \pi_F(x_{k_p}; \Omega_i)) \mid y \in \Omega\}$, we have

$$\begin{aligned}\sum_{i=1}^{m} \sigma_F(y - \pi_F(x_{k_p}; \Omega_i)) &\geq \sum_{i=1}^{m} \sigma_F(y_{k_p} - \pi_F(x_{k_p}; \Omega_i)) \geq \sigma_F(y_{k_p} - \pi_F(x_{k_p}; \Omega_1)) \\ &\geq \gamma \|y_{k_p} - \pi_F(x_{k_p}; \Omega_1)\|,\end{aligned}$$

where γ is the constant defined in Proposition 3.1. Letting $p \to \infty$, one obtains a contradiction showing that $\{y_k\}$ is bounded.

Now fix any subsequence $\{y_{k_\ell}\}$ of $\{y_k\}$ that converges to $\bar{y} \in \Omega$. Then

$$0 \in \sum_{i=1}^{m} \nabla \sigma_F(y_{k_\ell} - \pi(x_{k_\ell}; \Omega_i)) + N(y_{k_\ell}; \Omega),$$

which implies that

$$0 \in \sum_{i=1}^{m} \nabla \sigma_F(\bar{y} - \pi_F(\bar{x}; \Omega_i)) + N(\bar{y}; \Omega).$$

Therefore, $\bar{y} = \psi(\bar{x})$. It follows that $y_k = \psi(x_k)$ converges to $\bar{y} = \psi(\bar{x})$, so ψ is continuous at $\bar{x}$. Fix any $x \in \Omega$ such that $x \neq \psi(x)$. Since the function g is strictly convex on Ω and $\psi(x)$ is its unique minimizer on Ω, one has that

$$\begin{aligned}T(\psi(x)) = \sum_{i=1}^{m} d_F(\psi(x); \Omega_i) &\leq \sum_{i=1}^{m} \sigma_F(\psi(x) - \pi_F(x; \Omega_i)) \\ &< \sum_{i=1}^{m} \sigma_F(x - \pi_F(x; \Omega_i)) = T(x).\end{aligned}$$

The proof is now complete. ∎

Let us present below a convergence theorem for the MM algorithm.

THEOREM 4.9. *Consider the generalized Fermat–Torricelli problem* (11) *in the setting of Lemma* 4.8. *Let $\{x_k\}$ be a sequence generated by the MM algorithm, i.e.,*

$x_{k+1} = \psi(x_k)$ with a given $x_0 \in \Omega$. Then any subsequential limit of the sequence $\{x_k\}$ is an optimal solution of problem (11)*. If we assume additionally that Ω_i for $i = 1, \ldots, m$ are strictly convex, then $\{x_k\}$ converges to the unique optimal solution of the problem.*

Proof. In the setting of this theorem, [8, Proposition 1] implies that $\|x_{k+1} - x_k\| \to 0$. Since $x_{k+1} := \psi(x_k)$, applying Lemma 4.8 yields $T(x_{k+1}) \leq T(x_k) \leq \cdots \leq T(x_0)$ for every k. Then from the assumptions made, it is not hard to see that $\{x_k\}$ is a bounded sequence. Let $\{x_{k_\ell}\}$ be a subsequence of $\{x_k\}$ that converges to some $\bar{x}'$. Note that $\|x_{k+1} - x_k\| \to 0$ implies that $\{x_{k_\ell+1}\}$ also converges to $\bar{x}'$ as $\ell \to \infty$. Since $x_{k_\ell} \notin \Omega_i$ for all $i = 1, \ldots, m$ and for all ℓ, from the definition of the sequence $\{x_k\}$, one has

$$0 \in \sum_{i=1}^{m} \nabla \sigma_F(x_{k_\ell+1} - \pi_F(x_{k_\ell}; \Omega_i)) + N(x_{k_\ell+1}; \Omega).$$

Then

$$0 \in \sum_{i=1}^{m} \nabla \sigma_F(\bar{x}' - \pi_F(\bar{x}'; \Omega_i)) + N(\bar{x}'; \Omega).$$

Thus,

$$0 \in \sum_{i=1}^{m} \partial d_F(\bar{x}'; \Omega_i) + N(\bar{x}'; \Omega) = \partial T(\bar{x}') + N(\bar{x}'; \Omega).$$

Therefore, $\bar{x}'$ is an optimal solution of problem (11).

If Ω_i for $i = 1, \ldots, m$ are strictly convex, then problem (11) has a unique optimal solution $\bar{x}$ by Proposition 4.5. Thus, $\bar{x}' = \bar{x}$ and the original sequence $\{x_k\}$ converges to $\bar{x}$. ∎

It is important to note that the algorithm may not converge in general. Our examples partially answer the question raised in the concluding remarks of [8].

Example 4.10. Let Ω_1 and Ω_2 be subsets of $\mathbb{R}^2$ defined by

$$\Omega_1 := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geq 1\} \text{ and } \Omega_2 := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \leq -1\}.$$

Consider the generalized Fermat–Torricelli problem (11) for two sets Ω_1 and Ω_2 with the constraint being the line $\Omega := \mathbb{R} \times \{0\}$ generated by the ℓ_∞-norm, i.e., $F = \{(u_1, u_2) \in \mathbb{R}^2 \mid |u_1| + |u_2| \leq 1\}$. Starting from $x_0 = (0, 0)$, choose $y_{0,1} = (1, 1)$ and $y_{0,2} = (1, -1)$. Then $x_1 = (1, 0)$ is an optimal solution of the generalized Fermat–Torricelli problem for two points $y_{0,1}$ and $y_{0,2}$ generated by the ℓ_∞-norm. Similarly, we can choose $y_{1,1} = (2, 1)$, $y_{1,2} = (2, -1)$, and $x_2 = (2, 0)$. Repeating this process, one sees that $x_k = (k, 0)$ is a sequence generated by the MM algorithm, which does not have any convergent subsequence.

Example 4.11. Let Ω_1 and Ω_2 be subsets of $\mathbb{R}^2$ defined by

$$\Omega_1 := \{(x_1, 0) \in \mathbb{R}^2 \mid x_1 \geq 0\} \text{ and } \Omega_2 := \{(0, x_2) \in \mathbb{R}^2 \mid x_2 \geq 0\}.$$

Consider the unconstrained generalized Fermat–Torricelli problem (11) for two sets Ω_1 and Ω_2 generated by the ℓ_∞-norm. It is not hard to see that $(0, 0)$ is the unique optimal solution of this problem. Starting from $x_0 = (1/2, 1/2)$, choose $y_{0,1} = (1, 0)$ and $y_{0,2} = (0, 1)$. Then $x_1 = (1/2, 1/2)$ is an optimal solution of the generalized

Fermat–Torricelli problem for two points $y_{0,1}$ and $y_{0,2}$ generated by the ℓ_∞-norm. Obviously, if we choose the projections in this way, the sequence of optimal solutions and optimal values of the majorizations do not converge to the optimal solution and the optimal value of the problem.

Let us consider an example in which the convergence for the MM method is not guaranteed even if we consider generalized Fermat–Torricelli problems generated by the Euclidean norm.

Example 4.12. Let Ω_i for $i = 1, 2, 3$ be three Euclidean balls of $\mathbb{R}^2$ defined with radii 1 and centers at $(-4, 0)$, $(0, 0)$, and $(4, 0)$, respectively. Consider the unconstrained generalized Fermat–Torricelli problem (11) for these sets generated by the Euclidean norm. We use the starting point $x_0 = (0, 1)$. Then $y_{0,2} = x_0 = (0, 1)$, and $y_{0,1}$ and $y_{0,3}$ are the intersections of the line segments connecting x_0 with the centers of Ω_1 and Ω_3 and the boundaries of these disks. Obviously, $x_k = x_0$ for every k, where x_k is the sequence defined by the MM algorithm. However, x_0 is not an optimal solution of the problem. In fact, the solution set is the line segment connecting $(-1, 0)$ and $(1, 0)$.

Let Θ be a nonempty closed convex set. Consider the generalized distance function $d_F(\cdot; \Theta)$ generated by a dynamic F. For a point $\bar{x} \notin \Theta$, a point $\bar{w} \in \pi_F(\bar{x}; \Theta)$ is said to be a *representation* of the subdifferential $\partial d_F(\bar{x}; \Theta)$ if

$$\partial \sigma_F(\bar{x} - \bar{w}) \subseteq N(\bar{w}; \Theta).$$

From the definition we see that if F is normally smooth and round, then $\bar{w} := \pi_F(\bar{x}; \Theta)$ is always a representation of the subdifferential $\partial d_F(\bar{x}; \Theta)$.

Example 4.13. Let Θ be the cube $[c_1 - r, c_1 + r] \times [c_2 - r, c_2 + r] \times [c_3 - r, c_3 + r]$ of $\mathbb{R}^3$, and let

$$F := \{(u_1, u_2, u_3) \in \mathbb{R}^3 \mid |u_1| + |u_2| + |u_3| \le 1\}.$$

For any $x \notin \Theta$, the choice of projection

$$w := \{y \in \mathbb{R}^3 \mid y_i = \max\{c_i - r, \min\{x_i, c_i + r\}\}\} \in \pi_F(x; \Theta)$$

satisfies that condition that w is a representation of $\partial d_F(x; \Theta)$.

PROPOSITION 4.14. *Consider the generalized Fermat–Torricelli problem* (11). *Let* $\{x_k\}$ *be the sequence in the MM algorithm defined by*

$$x_{k+1} \in \arg\min \left\{ \sum_{i=1}^{m} \sigma_F(x - y_{k,i}) \mid x \in \Omega \right\},$$

where $y_{k,i} \in \pi_F(x_k; \Omega_i)$. *Suppose that* $\{x_k\}$ *converges to* $\bar{x}$ *that does not belong to* Ω_i *for* $i = 1, \ldots, m$. *Suppose further that for any limit point,* $\bar{y}_i \in \pi_F(\bar{x}; \Omega_i)$ *of* $\{y_{k,i}\}$ *is a representation of the subdifferential* $\partial d_F(\bar{x}; \Omega_i)$. *Then* $\bar{x}$ *is an optimal solution of the problem.*

Proof. For sufficiently large k, from the definition of the sequence $\{x_k\}$, one has

$$0 \in \sum_{i=1}^{m} \partial \sigma_F(x_{k+1} - y_{k,i}) + N(x_{k+1}; \Omega).$$

The estimate

$$\sigma_F(-y_{k,i}) \le \sigma_F(-x_k) + \sigma_F(x_k - y_{k,i}) \le \sup_k [\sigma_F(-x_k) + d_F(x_k; \Omega_i)] < \infty$$

implies that $\{y_{k,i}\}_k$ is a bounded sequence in Ω_i for $i = 1, \ldots, m$. Without loss of generality, we can assume that $y_{k,i} \to \bar{y}_i \in \pi_F(\bar{x}; \Omega_i)$ as $k \to \infty$. Using the fact that $\partial\sigma_F(u)$ is compact for any $u \in \mathbb{R}^n$ and the normal cone mapping $u \rightrightarrows N(u; \Omega)$ has closed graph yields

$$0 \in \sum_{i=1}^{m} \partial\sigma_F(\bar{x} - \bar{y}_i) + N(\bar{x}; \Omega).$$

Since $\partial\sigma_F(\bar{x} - \bar{y}_i) = \partial\sigma_F(\bar{x} - \bar{y}_i) \cap N(\bar{y}_i; \Omega_i) = \partial d_F(\bar{x}; \Omega_i)$,

$$0 \in \sum_{i=1}^{m} \partial d_F(\bar{x}; \Omega_i) + N(\bar{x}; \Omega) = \partial T(\bar{x}) + N(\bar{x}; \Omega).$$

Therefore, $\bar{x}$ is also an optimal solution of problem (11). □

Remark 4.15 (subgradient-type algorithms). The generalized Fermat–Torricelli problems presented in this section and section 3 can be solved by the projected subgradient method (see, e.g., [3, 29]). When applying the projected subgradient algorithm to the generalized Fermat–Torricelli problem (11), at iteration k we need to find a subgradient $u_{k,i}$ of each component function $\varphi_i(x) = d_F(x; \Omega_i)$ for $i = 1, \ldots, m$ at x_k. By the well-known *subdifferential sum rule* of convex analysis,

$$w_k := \sum_{i=1}^{m} u_{k,i}$$

is a subgradient of T at x_k. Proposition 4.3 as well as its specification to the case of the distance function in Corollary 4.4 provide us with a method of finding such a subgradient. Note that if $x_k \in \Omega_i$, the subdifferential $\partial d_F(x_k; \Omega_i)$ always contains 0, so we can choose $u_{k,i} = 0$. In the case where $x_k \notin \Omega_i$, a subgradient $u_{k,i}$ can be found by using a projection point $p_{k,i} \in \pi_F(x_k; \Omega_i)$ and we find $u_{k,i} \in \partial\sigma_F(x_k - p_{k,i}) \cap N(p_{k,i}; \Omega_i)$.

The projected subgradient algorithm exhibits slow convergence rates when applying to the generalized Fermat–Torricelli problems (9) and (11). One of the reasons is that in each iteration, in order to get an improvement we need to calculate all subgradients $u_{k,i}$ for $i = 1, \ldots, m$. This is computationally expensive if the number of target sets is large. In order to overcome this shortcoming, the stochastic subgradient method provides an alternative; see [3]. The main idea is that in each iteration, rather than scanning through all the target sets to find a subgradient as in the subgradient method, we choose t uniformly at random from I and find the subgradient $w_{k,t} \in \partial d_F(x_k; \Omega_t)$. After that, we define $\widetilde{w}_k := m w_{k,t}$ and perform the iteration

$$x_{k+1} := \pi(x_k - \alpha_k \widetilde{w}_k; \Omega).$$

A more general method can be presented as follows. Fix a positive integer p such that $p \leq |I|$. At the iteration k, we choose uniformly at random a nonempty set of indices I_k, $|I_k| = p$, that is a subset of I. Then for each $i \in I_k$, find $u_{k,i} \in \partial d_F(x_k; \Omega_i)$. After that, set

$$\widetilde{w}_k := m \frac{\sum_{i \in I_k} u_{k,i}}{p},$$

and perform the iteration

$$x_{k+1} := \pi(x_k - \alpha_k \widetilde{w}_k; \Omega),\ V_{k+1} := \min\{V_k, f(x_k)\}.$$

5. Numerical examples. To demonstrate the methods presented in the previous sections, let us consider a numerical example below.

Example 5.1. The latitude/longitude coordinates in decimal format of 1217 US cities are recorded, e.g., at http://www.realestate3d.com/gps/uslatlongdegmin.htm. We convert the longitudes provided by the website above from positive to negative to match with the real data. Our goal is to find a point that minimizes the sum of the distances to the given points representing the cities.

If we consider the case where $\sigma_F(x) = \|x\|$, the Euclidean norm, Algorithm 2 allows us to find an approximate optimal value $V^* \approx 23409.33$ and an approximate optimal solution $x^* \approx (38.63, -97.35)$. Similarly, if $\sigma_F(x) = \|x\|_1$, an approximate optimal value is $V^* \approx 28724.68$ and an approximate optimal solution is $x^* \approx (39.48, -97.22)$. With the same situation but considering the ℓ_∞-norm, an approximate optimal value is $V^* \approx 21987.76$ and an approximate optimal solution is $x^* \approx (37.54, -97.54)$.

Figure 2 below shows the relation between the number of iterations k and the optimal value $V_k = H(y_k)$ generated by different norms.

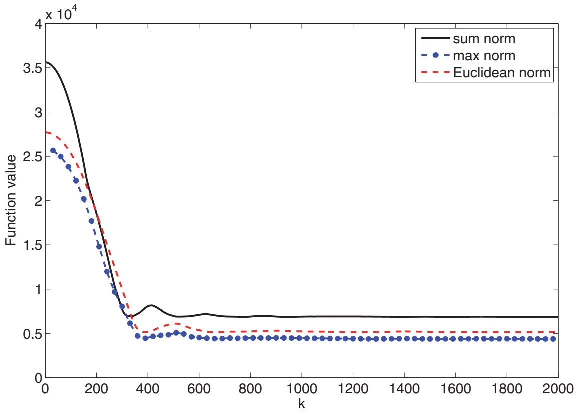

FIG. 2. *Generalized Fermat–Torricelli problems with different norms.*

In the example below we apply Algorithm 2 in combination with the Weiszfeld algorithm to solve generalized Fermat–Torricelli problems involving sets.

Example 5.2. In the same setting as Example 5.1, we consider 1217 squares centered at the coordinates of the cities with the same radius (half-side length) $r = 2$. The constraint is the line given by the equation $x - y = -180$. We implement Algorithm 3 with the starting point $x_0 = (0, 180)$ to solve the generalized Fermat–Torricelli problem generated by these squares and the Euclidean norm. In each step of the MM algorithm, we use Weiszfeld's algorithm to solve the classical Fermat–Torricelli problem generated by the projections $y_{k,i}$ for $i = 1, \ldots, 1217$. The MM method gives very fast convergence rate in this example. With 5 iterations of the MM algorithm along with 10 iterations of Weiszfeld's algorithm, we achieve an approximate optimal value $V^* \approx 38161.35$ and an approximate optimal solution $x^* \approx (56.84, -123.16)$; see Figure 3. It is required to perform more than $15,000$ iterations of the stochastic subgradient algorithm to achieve similar results. However, the MM algorithm may not converge in some situations where the sequence x_k enters the target sets, while the stochastic subgradient method is applicable to this case.

Example 5.3. Consider six given cubes in $\mathbb{R}^3$ with centers defined by the rows of

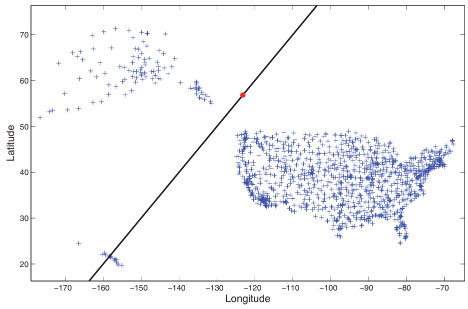

FIG. 3. *A generalized Fermat–Torricelli problem with US cities.*

the matrix

$$\begin{pmatrix} -6 & 6 & -4 \\ -5 & -3 & -6 \\ 2 & 3 & 4 \\ 4 & -4 & -5 \\ 5 & 6 & -6 \\ -5 & -2 & 4 \end{pmatrix}$$

and the half-side lengths being $r_i = 1.5$ for $i = 1, \ldots, 6$. The implementation of the algorithm above for the generalized Fermat–Torricelli problem for the cubes generated by the Euclidean norm yields a suboptimal solution $x^* = (-1.0405, 0.8402, -1.4322)$. This result can also be obtained by the subgradient method under a much slower convergence rate.

With the same problem but considering the ℓ_∞-norm instead of the Euclidean norm, the choice of the projection from a point x to any cube Ω with center c and half-side length r is given by

$$\{y \in \mathbb{R}^3 \mid y_i = \max\{c_i - r, \min\{x_i, c_i + r\}\}\} \in \pi_F(x; \Omega).$$

Then one obtains a suboptimal solution $x^* = (-0.6511, 0.6511, -0.3489)$; see Figure 4.

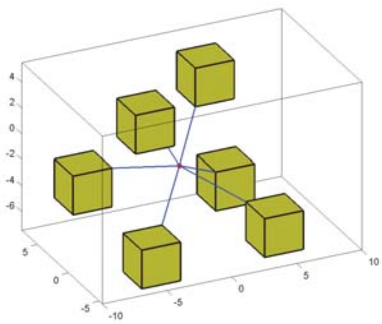

FIG. 4. *A generalized Fermat–Torricelli problem with the MM method.*

Acknowledgment. We are grateful to the anonymous referees for their valuable suggestions and remarks that allowed us to improve the original presentation.

REFERENCES

[1] H. Avron, S. Kale, S. Kasiviswanathan, and V. Sindhwani, *Efficient and practical stochastic subgradient descent for nuclear norm regularization*, in Proceedings of the ACM International Conference on Machine Learning (ICML), 2012.

[2] M. Bacak, *Computing Medians and Means in Hadamard Spaces*, preprint, arXiv:1210.2145, 2012.

[3] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Boston, MA, 2003.

[4] R. I. Boţ and C. Hendrich, *A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM J. Optim., 23 (2013), pp. 2541–2565.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[6] J. V. Burke, M. C. Ferris, and M. Qian, *On the Clarke subdifferential of the distance function of a closed set*, J. Math. Anal. Appl., 166 (1992), pp. 199–213.

[7] E. Chi and K. Lange, *A look at the generalized Heron problem through the lens of majorization-minimization*, Amer. Math. Monthly, 121 (2014), pp. 95–108.

[8] E. Chi, H. Zhou, and K. Lange, *Distance majorization and its applications*, Math. Program., 146 (2014), pp. 409–436.

[9] F. H. Clarke, *Nonsmooth Analysis and Optimization*, John Wiley & Sons, New York, 1983.

[10] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, *Efficient projections onto the ℓ_1-ball for learning in high dimensions*, in Proceedings of the 25th International ACM Conference on Machine Learning, 2008, pp. 272–279.

[11] 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.

[12] D. R. Hunter and K. Lange, *Tutorial on MM algorithms*, Amer. Statist., 58 (2004), pp. 30–37.

[13] K.W. Kuhn, *"Steiner's" problem revisited*, in Studies in Optimization, Stud. Math. 10, Math Association of America, Washington, D.C., 1974, pp. 52–70.

[14] Y. S. Kupitz and H. Martini, *Geometric aspects of the generalized Fermat-Torricelli problem*, Bolyai Soc. Math. Stud., 6 (1997), pp. 55–127.

[15] Y. S. Kupitz and H. Martini, *The Fermat-Torricelli problem, Part* I: *A discrete gradient-method approach*, J. Optim. Theory Appl., 158 (2013), pp. 305–327.

[16] K. Lange, D. R. Hunter, and I. Yang, *Optimization transfer using surrogate objective functions (with discussion)*, J. Comput. Graph. Statist., 9 (2000), pp. 1–59.

[17] H. Martini, K. J. Swanepoel, and G. Weiss, *The Fermat-Torricelli problem in normed planes and spaces*, J. Optim. Theory Appl., 115 (2002), pp. 283–314.

[18] B. S. Mordukhovich and N. M. Nam, *An Easy Path to Convex Analysis and Applications*, Synth. Lect. Math. Stat. 14, S. G. Krantz, ed., Morgan & Claypool Publishers, Williston, VT, 2014.

[19] B. S. Mordukhovich and N. M. Nam, *Applications of variational analysis to a generalized Fermat-Torricelli problem.* J. Optim. Theory Appl., 148 (2011), pp. 431–454.

[20] B. S. Mordukhovich, N. M. Nam, and J. Salinas, *Applications of variational analysis to a generalized Heron problem*, Appl. Anal., 91 (2012), pp. 1915–1942.

[21] N. M. Nam and N. Hoang, *A generalized Sylvester problem and a generalized Fermat-Torricelli problem*, J. Convex Anal., 20 (2013), pp. 669–687.

[22] A. Nedić and D. P. Bertsekas, *Incremental subgradient methods for nondifferentiable optimization*, SIAM J. Optim., 12 (2001), pp. 109–138.

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

[24] Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course*, Appl. Optim. 87, Kluwer Academic Publishers, Boston, MA, 2004.

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

[26] S. Nickel, J. Puerto, and A. M. Rodriguez-Chia, *An approach to location models involving sets as existing facilities*, Math. Oper. Res., 28 (2003), pp. 693–715.

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

[28] S. Shalev-Shwartz, Y. Singer, and N. Srebro, *PEGASOS: Primal Estimated sub-GrAdient SOlver for SVM*, in Proceedings of the ACM International Conference on Machine Learning (ICML), 2007.
[29] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer-Verlag, Berlin, 1985.
[30] E. Weiszfeld, *Sur le point pour lequel la somme des distances de n points donnés est minimum*, Tohoku Math. J., 43 (1937), pp. 355–386.
[31] E. Weiszfeld and F. Plastria, *On the point for which the sum of the distances to n given points is minimum*. Ann. Oper. Res., 167 (2009), pp. 7–41.
[32] T. Zhou, D. Tao, and X. Wu, *NESVM, a fast gradient method for support vector machines*, in Proceedings of the IEEE International Conference on Data Mining (ICDM), 2010.