SIAM J. OPTIM.
Vol. 11, No. 2, pp. 389–410

A SMOOTHING NEWTON METHOD FOR MINIMIZING A SUM OF EUCLIDEAN NORMS*

LIQUN QI†‡ AND GUANGLU ZHOU†

Abstract. We consider the problem of minimizing a sum of Euclidean norms, $f(x) = \sum_{i=1}^m \|b_i - A_i^T x\|$. This problem is a nonsmooth problem because f is not differentiable at a point x when one of the norms is zero. In this paper we present a smoothing Newton method for this problem by applying the smoothing Newton method proposed by Qi, Sun, and Zhou [*Math. Programming*, 87 (2000), pp. 1–35] directly to a system of strongly semismooth equations derived from primal and dual feasibility and a complementarity condition. This method is globally and quadratically convergent. As applications to this problem, smoothing Newton methods are presented for the Euclidean facilities location problem and the Steiner minimal tree problem under a given topology. Preliminary numerical results indicate that this method is extremely promising.

Key words. sum of norms, smoothing Newton method, semismoothness, Euclidean facilities location, shortest networks, Steiner minimum trees

AMS subject classifications. 90C33, 90C30, 65H10

PII. S105262349834895X

1. Introduction.

Consider the problem of minimizing a sum of Euclidean norms (MSNs):

$$\min_{x\in R^n} \sum_{i=1}^m \|b_i - A_i^T x\|, \tag{1.1}$$

where $b_1, b_2, \ldots, b_m \in R^d$ are column vectors in the Euclidean d-space, $A_1, A_2, \ldots, A_m \in R^{n\times d}$ are $n \times d$ matrices with each having full column rank, $n \le m(d-1)$, and $\|r\|$ represents the Euclidean norm $(\sum_{i=1}^m r_i^2)^{1/2}$. Let $A = [A_1, A_2, \ldots, A_m]$. In what follows we always assume that A has rank n. Let

$$f(x) = \sum_{i=1}^m \|b_i - A_i^T x\|. \tag{1.2}$$

It is clear that $x = 0$ is an optimal solution to problem (1.1) when all of the b_i are zero. Therefore, we assume in the rest of this paper that not all of the b_i are zero. Problem (1.1) is a convex programming problem, but its objective function f is not differentiable at any point x when some $b_i - A_i^T x = 0$. Three special cases of this problem are the Euclidean single facility location (ESFL) problem, the Euclidean multifacility location (EMFL) problem, and the Steiner minimal tree (SMT) problem under a given topology.

Many algorithms have been designed to solve problem (1.1). For the ESFL problem, Weiszfeld [34] gave a simple iterative algorithm in 1937. Later, a number of

*Received by the editors November 30, 1998; accepted for publication (in revised form) March 23, 2000; published electronically October 18, 2000. This work was supported by the Australian Research Council.

http://www.siam.org/journals/siopt/11-2/34895.html

†School of Mathematics, The University of New South Wales, Sydney 2052, Australia (zhou@maths.unsw.edu.au).

‡Current address: Department of Applied Mathematics, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (maqilq@polyu.edu.hk).

important results were obtained along this line; see [6, 7, 13, 22, 24, 30, 31, 33]. Practical algorithms for solving these problems began with the work of Calamai and Conn [4, 5] and Overton [25], where they proposed projected Newton algorithms with the quadratic rate of convergence. The essential idea of these algorithms is as follows. In each iteration a search direction is computed by Newton's method projected into a linear manifold along which f is locally differentiable. The advantage of this method is the quadratic convergence and the avoidance of approximation techniques for f. However, it is difficult to use this method due to the dynamic structure of the linear manifold into which the method projects the search direction. Every time terms are added and deleted from the active set, the size and the sparse structure of the problem changes.

More recently, Andersen [1] used the HAP idea [13] to smooth the objective function by introducing a perturbation $\varepsilon > 0$ and applied a Newton barrier method for solving this problem. Andersen et al. [3] proposed a primal-dual interior-point method based on the ε-perturbation and presented impressive computational results. Xue and Ye [35, 36] presented polynomial-time primal-dual potential reduction algorithms by transforming this problem into a standard convex programming problem in conic form. However, these methods do not possess second-order convergence.

In recent years, two major reformulation approaches, the nonsmooth approach and the smoothing approach, for solving nonlinear complementarity problems (NCPs) and box constrained variational inequality problems (BVIPs), have been rapidly developed based on NCP and BVIP functions, e.g., see [8, 9, 10, 11, 14, 15, 19, 21, 26, 28, 32, 37, 38] and references therein. In particular, Jiang and Qi [21] and De Luca, Facchinei, and Kanzow [14] proposed globally and superlinearly (quadratically) convergent nonsmooth Newton methods for NCPs, which only require solving a system of linear equations to determine the search direction at each iteration. A globally and superlinearly (quadratically) convergent smoothing Newton method was proposed by Chen, Qi, and Sun in [10], where the authors exploited a Jacobian consistence property and applied this property to an infinite sequence of smoothing approximation functions to get high-order convergence. On the other hand, Hotta and Yoshise [20], Qi, Sun, and Zhou [28], and Jiang [19] proposed smoothing methods for NCPs and BVIPs by treating the smoothing parameter as a variable, in which the smoothing parameter is driven to zero automatically and no additional procedure for adjusting the smoothing parameter is necessary. Some regularized versions of the method in [28] were proposed in [26, 32, 38] for NCPs and BVIPs.

In this paper we present a smoothing Newton method for problem (1.1) by applying the smoothing Newton method proposed by Qi, Sun, and Zhou [28] directly to a system of strongly semismooth equations derived from primal and dual feasibility and a complementarity condition and prove that this method is globally and quadratically convergent. Numerical results indicate that this method is extremely promising.

This paper is organized as follows. In section 2, we transform primal and dual feasibility and a complementarity condition derived from problem (1.1) and its dual problem into a system of strongly semismooth equations. Some smooth approximations to the projection operator on the unit ball are given in section 3. In section 4, we present a smoothing Newton method for solving problem (1.1) and prove that this method is globally and quadratically convergent. In section 5, we discuss applications to the ESFL problem, the EMFL problem, and the SMT problem. In section 6, we present some numerical results. We conclude this paper in section 7.

Concerning notation, for a continuously differentiable function $F : R^n \to R^m$, we denote the Jacobian of F at $x \in R^n$ by $F'(x)$, whereas the transposed Jacobian is denoted as $\nabla F(x)$. In particular, if $m = 1$, the gradient $\nabla F(x)$ is viewed as a column vector.

Let $F : R^n \to R^m$ be a locally Lipschitzian vector function. By Rademacher's theorem, F is differentiable almost everywhere. Let Ω_F denote the set of points where F is differentiable. Then the B-subdifferential of F at $x \in R^n$ is defined as

$$\partial_B F(x) = \{ \lim_{\substack{x^k \to x \\ x^k \in \Omega_F}} \nabla F(x^k)^T \}, \tag{1.3}$$

while Clarke's generalized Jacobian of F at x is defined as

$$\partial F(x) = \mathrm{conv}\partial_B F(x), \tag{1.4}$$

(see [12, 27, 29]). F is called *semismooth* at x if F is directionally differentiable at x and for all $V \in \partial F(x + h)$ and $h \to 0$,

$$F'(x; h) = Vh + o(\|h\|); \tag{1.5}$$

F is called *p-order semismooth*, $p \in (0, 1]$, at x if F is semismooth at x and for all $V \in \partial F(x + h)$ and $h \to 0$,

$$F'(x; h) = Vh + O(\|h\|^{1+p}); \tag{1.6}$$

F is called *strongly semismooth* at x if F is 1-order semismooth at x. F is called a (strongly) semismooth function if it is (strongly) semismooth everywhere (see [27, 29]). In particular, a PC^2 (piecewise twice continuously differentiable) function is a strongly semismooth function. Here, $o(\|h\|)$ stands for a vector function $e : R^n \to R^m$, satisfying

$$\lim_{h \to 0} \frac{e(h)}{\|h\|} = 0,$$

while $O(\|h\|^2)$ stands for a vector function $e : R^n \to R^m$, satisfying

$$\|e(h)\| \le M\|h\|^2$$

for all h satisfying $\|h\| \le \delta$ and some $M > 0$ and $\delta > 0$.

LEMMA 1.1 (see [29]).

(i) *If F is semismooth at x, then for any $h \to 0$,*

$$F(x + h) - F(x) - F'(x; h) = o(\|h\|);$$

(ii) *if F is p-order semismooth at x, then for any $h \to 0$,*

$$F(x + h) - F(x) - F'(x; h) = O(\|h\|^{1+p}).$$

THEOREM 1.2 (see [16, Theorem 19]). *Suppose that the function $\mathcal{F} : R^n \to R^m$ is p-order semismooth at x and the function $\mathcal{G} : R^m \to R^l$ is p-order semismooth at $\mathcal{F}(x)$. Then the composite function $\mathcal{H} = \mathcal{G} \circ \mathcal{F}$ is p-order semismooth at x.*

For a set $\mathcal{A}$, $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$. We denote $x^T x$ by x^2, for a vector $x \in R^n$, i.e., $x^2 = \|x\|^2$. For $A \in R^{n\times m}$, $\|A\|$ denotes the induced norm, i.e., $\|A\| = \max\{\|Au\| : u \in R^n, \|u\| = 1\}$. Let I_d denote the $d \times d$ identity matrix. Let $b^T = [b_1^T, \dots, b_m^T]$, $y = [y_1^T, \dots, y_m^T]^T \in R^{md}$, $R_+ = \{\varepsilon \in R : \varepsilon \ge 0\}$, and $R_{++} = \{\varepsilon \in R : \varepsilon > 0\}$. Finally, we use $\varepsilon \downarrow 0^+$ to denote the case that a positive scalar ε tends to 0.

2. Some preliminaries. In [1, 3], Andersen et al. studied the duality for problem (1.1) and presented some efficient algorithms for solving it. In this section we will transform three sets of equations—primal feasibility, dual feasibility, and the complementarity condition derived from problem (1.1) and its dual problem—into a system of strongly semismooth equations. This transformation is very important for the method proposed in this paper.

LEMMA 2.1. *Assume that A has rank n. Then the set of solutions to the problem* (1.1) *is bounded.*

Proof. It follows from the assumed rank of A that

$$\min_{\|x\|=1} \|A^T x\| = \tau > 0. \tag{2.1}$$

From (2.1) we obtain

$$\|A^T x\| \geq \tau \|x\|. \tag{2.2}$$

This shows that the set of solutions to the problem (1.1) is bounded. ∎

The dual of the problem (1.1) has the form (see [1])

$$\max_{y \in Y} b^T y, \tag{2.3}$$

where

$$Y = \left\{ y = [y_1^T, \dots, y_m^T]^T \in R^{md} \ : \ y_i \in R^d, \ \|y_i\| \leq 1, \ i = 1, \dots, m; Ay = 0 \right\}. \tag{2.4}$$

THEOREM 2.2 (see [1]). *Let $x \in R^n$, $y \in Y$ and let $x^* \in R^n$, $y^* \in Y$ be optimal solutions to problems* (1.1) *and* (2.3), *respectively. Then*

$$(a)\ b^T y \leq \sum_{i=1}^m \|b_i - A_i^T x\| \quad \text{(weak duality)}$$

and

$$(b)\ b^T y^* = \sum_{i=1}^m \|b_i - A_i^T x^*\| \quad \text{(strong duality)}.$$

DEFINITION 2.3 (see [1]). *A solution $x \in R^n$ and a solution $y \in Y$ are called ε-optimal to problems* (1.1) *and* (2.3) *if*

$$\sum_{i=1}^m \|b_i - A_i^T x\| - b^T y \leq \varepsilon.$$

From Theorem 2.2 we have that (x^*, y^*) is a pair of optimal solutions to problems (1.1) and (2.3) if and only if (x^*, y^*) is a solution to the following system:

$$\begin{cases} Ay = 0, \\ \|y_i\| \leq 1, \ i = 1, \dots, m, \\ \displaystyle\sum_{i=1}^m \|b_i - A_i^T x\| - b^T y = 0. \end{cases} \tag{2.5}$$

Suppose that $y \in R^{md}$, satisfying that $Ay = 0$ and $\|y_i\| \leq 1, i = 1, 2, \ldots, m$. Then

$$\begin{aligned}\sum_{i=1}^m \|b_i - A_i^T x\| - b^T y &= \sum_{i=1}^m \|b_i - A_i^T x\| - \sum_{i=1}^m b_i^T y_i \\ &= \sum_{i=1}^m \left(\|b_i - A_i^T x\| - b_i^T y_i\right) \\ &= \sum_{i=1}^m \left(\|b_i - A_i^T x\| - (b_i - A_i^T x)^T y_i + x^T(A_i y_i)\right) \\ &= \sum_{i=1}^m \left(\|b_i - A_i^T x\| - (b_i - A_i^T x)^T y_i\right) + x^T(Ay) \\ &= \sum_{i=1}^m \left(\|b_i - A_i^T x\| - (b_i - A_i^T x)^T y_i\right),\end{aligned}$$

and for $i = 1, 2, \ldots, m$,

$$\|b_i - A_i^T x\| - (b_i - A_i^T x)^T y_i \geq 0.$$

So the duality gap is zero if and only if

$$\|b_i - A_i^T x\| - (b_i - A_i^T x)^T y_i = 0$$

for $i = 1, \ldots, m$. Then (2.5) is equivalent to

$$\left\{\begin{array}{l} Ay = 0, \\ \|y_i\| \leq 1,\ i = 1, \ldots, m, \\ \|b_i - A_i^T x\| - (b_i - A_i^T x)^T y_i = 0,\ i = 1, \ldots, m. \end{array}\right. \tag{2.6}$$

LEMMA 2.4. *Let $r, s \in R^d$. If $\|s\| \leq 1$, then $\|r\| = r^T s$ if and only if $r - \|r\|s = 0$.*

Proof. Suppose $\|r\| = r^T s$. If $r = 0$, then $r - \|r\|s = 0$. If $r \neq 0$, then

$$\|r\| = r^T s \leq \|r\|\|s\|.$$

So $\|s\| = 1$. Then $(r - \|r\|s)^2 = \|r\|^2 - 2\|r\|r^T s + \|r\|^2\|s\|^2 = 0$, i.e., $r - \|r\|s = 0$.

On the other hand, if $r = 0$, then $\|r\| = r^T s$. If $r - \|r\|s = 0$ and $r \neq 0$, then $\|s\| = 1$ and $r^T s - \|r\|s^T s = r^T s - \|r\| = 0$, i.e., $\|r\| = r^T s$. ∎

From the above lemma (2.6) is equivalent to

$$\left\{\begin{array}{l} Ay = 0, \\ \\ \|y_i\| \leq 1,\ i = 1, \ldots, m, \\ \\ (b_i - A_i^T x) - \|b_i - A_i^T x\| y_i = 0,\ i = 1, \ldots, m. \end{array}\right. \tag{2.7}$$

It follows from (2.7) that if (x^*, y^*) is a pair of optimal solutions to problems (1.1) and (2.3), then for $i = 1, \ldots, m$, either $b_i - A_i^T x^* = 0$ or $\|y_i^*\| = 1$. We say strict complementarity holds at (x^*, y^*) if, for each i, only one of these two conditions holds.

Let $B = \{s \in R^d \colon \|s\| \leq 1\}$ and let $\Pi_B(s)$ be the projection operator onto B.

LEMMA 2.5. *Let $r, s \in R^d$. Then $s = \Pi_B(s + r)$ if and only if $\|s\| \leq 1$ and $\|r\| = r^T s$.*

Proof. Suppose that $s = \Pi_B(s+r)$. Then $\|s\| \le 1$ and

$$r^T(s - s^*) \ge 0 \text{ for any } \; s^* \in B.$$

It follows that $\|r\| = \max_{\|s^*\|\le 1} r^T s^* \le r^T s$. So $\|r\| = r^T s$.

On the other hand, if $\|r\| = r^T s$ and $\|s\| \le 1$, then for any $s^* \in B$,

$$r^T(s - s^*) \ge 0$$

because $\|r\| = \max_{\|s^*\|\le 1} r^T s^*$. Hence $s = \Pi_B(s+r)$. ◻

It follows from the above lemma that (2.6) is equivalent to

$$\text{(2.8)} \qquad \begin{cases} Ay = 0, \\ \\ y_i - \Pi_B(y_i + b_i - A_i^T x) = 0, \;\; i = 1, \dots, m. \end{cases}$$

Define $F : R^{n+md} \to R^{n+md}$ by

$$\text{(2.9)} \qquad \begin{cases} F_j(x,y) = (Ay)_j, \;\; j = 1, \dots, n, \\ \\ F_j(x,y) = y_i - \Pi_B(y_i + b_i - A_i^T x), \\ \qquad\qquad j = n + il, \quad i = 1, \dots, m, \quad l = 1, \dots, d. \end{cases}$$

Then we have that (x^*, y^*) is a pair of optimal solutions to problems (1.1) and (2.3) if and only if (x^*, y^*) is a solution to the following equation:

$$\text{(2.10)} \qquad F(x,y) = 0.$$

From Lemma 2.1, (2.3), and (2.4), we have the following.

LEMMA 2.6. *All solutions to* (2.10) *are bounded.*

Clearly, F is not continuously differentiable, but we can prove that it is strongly semismooth.

THEOREM 2.7. *The function F defined in* (2.9) *is strongly semismooth on $R^n \times R^{md}$.*

Proof.

$$\Pi_B(s) = \begin{cases} \dfrac{s}{\|s\|} & \text{if } \|s\| > 1, \\ s & \text{if } \|s\| \le 1. \end{cases}$$

Then

$$\text{(2.11)} \qquad \Pi_B(s) = \frac{s}{\max\{1, \|s\|\}} = \frac{s}{1 + \max\{0, (\|s\| - 1)\}}.$$

Since the function h, defined by $h(x) = \|x\|$, where $x \in R^d$, max functions, and linear functions are all strongly semismooth, from Theorem 1.2 F is strongly semismooth on $R^n \times R^{md}$. ◻

3. Smooth approximations to $\Pi_B(s)$.

In this section we will present some smooth approximations to the projection operator $\Pi_B(s)$ and study the properties of these smooth approximations.

In [9], Chen and Mangasarian presented a class of smooth approximations to the function $\max\{0, \cdot\}$. Similarly, we can give a class of smooth approximations to the projection operator $\Pi_B(s)$ defined in (2.11). For simplicity, throughout this paper we use only the following smooth function to approximate $\Pi_B(s)$, which is based on the neural networks smooth function and defined as follows:

$$\phi(t,s) = \frac{s}{q(t,s)},\ (t,s) \in R_{++} \times R^d, \tag{3.1}$$

where $q(t,s) = t\ln(e^{\frac{1}{t}} + e^{\frac{\sqrt{\|s\|^2+t^2}}{t}})$.

PROPOSITION 3.1. *$\phi(t,s)$ has the following properties:*

(i) *For any given $t > 0$, $\phi(t,s)$ is continuously differentiable;*

(ii) *$\phi(t,s) \in \mathrm{int}B$, for any given $t > 0$;*

(iii) *$|\phi(t,s) - \Pi_B(s)| \leq (\ln 2 + 1)t$;*

(iv) *for any given $t > 0$,*

$$\nabla\phi_s(t,s) = \frac{1}{q(t,s)} I_d - \frac{ss^T}{q(t,s)^2(1+e^{(1-\sqrt{\|s\|^2+t^2})/t})\sqrt{\|s\|^2+t^2}}, \tag{3.2}$$

and $\nabla\phi_s(t,s)$ is symmetric, positive definite and $\|\nabla\phi_s(t,s)\| < 1$;

(v) *for any given $s \in R^d$ and $t > 0$,*

$$\nabla\phi_t(t,s) = -\frac{1}{q^2(t,s)}\left(\ln e(t,s) - \frac{e^{\frac{1}{t}}}{te(t,s)} + \frac{\|s\|^2 e^{\frac{\sqrt{\|s\|^2+t^2}}{t}}}{t\sqrt{\|s\|^2+t^2}\,e(t,s)}\right)s, \tag{3.3}$$

where $e(t,s) = e^{\frac{1}{t}} + e^{\frac{\sqrt{\|s\|^2+t^2}}{t}}$.

Proof. It is clear that (i) holds. For any $t > 0$, $q(t,s) > \max\{1, \|s\|\}$. So (ii) holds. By Proposition 2.2(ii) in [9],

$$|q(t,s) - \max\{1,\|s\|\}| \leq (\ln 2 + 1)t.$$

Hence,

$$\begin{aligned} \|\phi(t,s) - \Pi_B(s)\| &= \frac{\|s\||q(t,s) - \max\{1,\|s\|\}|}{q(t,s)\max\{1,\|s\|\}} \\ &\leq |q(t,s) - \max\{1,\|s\|\}| \\ &\leq (\ln 2 + 1)t. \end{aligned}$$

By simple computation, (iv) and (v) hold. ∎

Let

$$p(t,s) = \begin{cases} \phi(|t|,s) & \text{if } t \neq 0, \\ \Pi_B(s) & \text{if } t = 0. \end{cases} \tag{3.4}$$

From Proposition 3.1 of [28] and Theorem 1.2 we have the following.

PROPOSITION 3.2. *$p(t,s)$ is a strongly semismooth function on $R \times R^d$.*

It follows from Proposition 3.1 that the following proposition holds.

PROPOSITION 3.3.

(i) *If $\|s^*\| < 1$, then*

$$\lim_{\substack{t^k \downarrow 0+ \\ s^k \to s^*}} \nabla\phi_s(t^k, s^k) = I_d;$$

(ii) *if* $\|s^*\| > 1$, *then*

$$\lim_{\substack{t^k \downarrow 0^+ \\ s^k \to s^*}} \nabla \phi_s(t^k, s^k) = \frac{1}{\|s^*\|} I_d - \frac{1}{\|s^*\|^3} s^*(s^*)^T,$$

which is symmetric, nonnegative definite, and the norm of this matrix is less than 1 *and the rank of this matrix is* $d - 1$.

4. A smoothing Newton method. In this section we will present a smoothing Newton method for solving problem (1.1) by applying the smoothing Newton method proposed by Qi, Sun, and Zhou [28] directly to the system of strongly semismooth equation (2.10) and prove that this method is globally and quadratically convergent.

Define $G : R \times R^n \times R^{md} \to R^{n+md}$ by

$$\left\{ \begin{array}{l} G_j(t,x,y) = (Ay)_j - tx_j, \ \ j = 1, \dots, n, \\ \\ G_j(t,x,y) = (y_i)_l - (p(t, y_i + b_i - A_i^T x))_l, \\ \qquad\qquad j = n + il, \quad i = 1, \dots, m, \quad l = 1, \dots, d. \end{array} \right. \tag{4.1}$$

Then G is continuously differentiable at any (t, x, y) with $t \neq 0$ and from Theorem 1.2 and Proposition 3.2 it is strongly semismooth on $R \times R^n \times R^{md}$.

Let $z := (t, x, y) \in R \times R^n \times R^{md}$ and define $H : R \times R^n \times R^{md} \to R^{n+md+1}$ by

$$H(z) := \begin{pmatrix} t \\ G(z) \end{pmatrix}. \tag{4.2}$$

Then H is continuously differentiable at any $z \in R_{++} \times R^n \times R^{md}$ and strongly semismooth at any $z \in R \times R^n \times R^{md}$, and $H(t^*, x^*, y^*) = 0$ if and only if $t^* = 0$ and $F(x^*, y^*) = 0$.

Let $p(t, y + b - A^T x) = [p(t, y_1 + b_1 - A_1^T x)^T, \dots, p(t, y_m + b_m - A_m^T x)^T]^T$.

LEMMA 4.1. *For any* $z = (t, x, y) \in R_{++} \times R^n \times R^{md}$,

$$H'(z) := \begin{pmatrix} 1 & 0 & 0 \\ -x & -tI_n & A \\ E(z) & P(z)A^T & I_{md} - P(z) \end{pmatrix}, \tag{4.3}$$

where

$$E(z) = \nabla p_t(t, y + b - A^T x), \tag{4.4}$$

and

$$P(z) = \mathrm{Diag}(p'_s(t, y_i + b_i - A_i^T x)), \tag{4.5}$$

and $H'(z)$ *is nonsingular.*

Proof. We have that (4.3) holds by simple computation. For any $z = (t, x, y) \in R_{++} \times R^n \times R^{md}$, in order to prove $H'(z)$ is nonsingular, we need to prove only that

$$M = \begin{pmatrix} -tI_n & A \\ P(z)A^T & I_{md} - P(z) \end{pmatrix}$$

is nonsingular. For any $t > 0$ and $(x, y) \in R^n \times R^{md}$, from Proposition 3.1 $P(z)$ is symmetric positive definite and $\|P(z)\| < 1$. Let $Mg = 0$, where $g = (g_1^T, g_2^T)^T \in R^n \times R^{md}$. Then we have

$$-tI_n g_1 + Ag_2 = 0, \tag{4.6}$$

and

$$P(z)A^T g_1 + (I_{md} - P(z))g_2 = 0. \tag{4.7}$$

From (4.7) we have

$$g_2 = -(I_{md} - P(z))^{-1} P(z) A^T g_1. \tag{4.8}$$

Then

$$(tI_n + A(I_{md} - P(z))^{-1} P(z) A^T) g_1 = 0. \tag{4.9}$$

Let

$$B(z) = tI_n + A(I_{md} - P(z))^{-1} P(z) A^T. \tag{4.10}$$

Then $B(z)$ is an $n \times n$ symmetric positive definite matrix because A has full rank. So $g_1 = 0$. Thus $g = 0$. This implies that M is nonsingular. So $H'(z)$ is nonsingular. ∎

Choose $\bar{t} \in R_{++}$ and $\gamma \in (0,1)$ such that $\gamma\bar{t} < 1$. Let $\bar{z} := (\bar{t}, 0, 0) \in R \times R^n \times R^{md}$. Define the merit function $\psi : R \times R^n \times R^{md} \to R_+$ by

$$\psi(z) := \|H(z)\|^2.$$

ψ is continuously differentiable on $R_{++} \times R^n \times R^{md}$ and strongly semismooth on $R \times R^n \times R^{md}$. Define $\beta : R_+ \times R^n \times R^{md} \to R_+$ by

$$\beta(z) := \gamma \min\{1, \psi(z)\}.$$

Let

$$\Omega := \{z = (t, x, y) \in R \times R^n \times R^{md} \mid t \geq \beta(z)\bar{t}\ \}.$$

Then, because for any $z \in R \times R^n \times R^{md}$, $\beta(z) \leq \gamma < 1$, it follows that for any $(x, y) \in R^n \times R^{md}$,

$$(\bar{t}, x, y) \in \Omega.$$

ALGORITHM 4.1.

Step 0. *Choose constants* $\delta \in (0,1)$ *and* $\sigma \in (0, 1/2)$. *Let* $z^0 := (\bar{t}, x^0, y^0) \in R_{++} \times R^n \times R^{md}$ *and* $k := 0$.

Step 1. *If* $H(z^k) = 0$, *then stop. Otherwise, let* $\beta_k := \beta(z^k)$.

Step 2. *Compute* $\Delta z^k := (\Delta t^k, \Delta x^k, \Delta y^k) \in R \times R^n \times R^{md}$ *by*

$$H(z^k) + H'(z^k)\Delta z^k = \beta_k \bar{z}. \tag{4.11}$$

Step 3. *Let* j_k *be the smallest nonnegative integer* j *satisfying*

$$\psi(z^k + \delta^j \Delta z^k) \leq [1 - 2\sigma(1 - \gamma\bar{t}\ )\delta^j]\psi(z^k). \tag{4.12}$$

Define $z^{k+1} := z^k + \delta^{j_k} \Delta z^k$.

Step 4. *Replace* k *by* $k+1$ *and go to Step* 1.

Remark. We can solve (4.11) in the following way: Let $\Delta t^k = -t^k + \beta_k \bar{t}$. Solve

$$B(z^k)\Delta x^k = -A(I_{md} - P(z^k))^{-1}(y^k - p^k + \Delta t^k E(z^k)) + (Ay^k - (t^k + \Delta t^k)x^k) \tag{4.13}$$

to get Δx^k, where $B(z^k)$ is defined in (4.10) and $p^k = p(t^k, y^k + b - A^T x^k)$. Then

$$\Delta y^k = -(I_{md} - P(z^k))^{-1} P(z^k) A^T \Delta x^k - (I_{md} - P(z^k))^{-1}(y^k - p^k + \Delta t^k E(z^k)).$$

Equation (4.13) is an n-dimensional symmetric positive definite linear system.

From Proposition 4.5 of [28] and Lemma 4.1 of [32] we have the following.

PROPOSITION 4.2. *Algorithm* 4.1 *is well defined at the kth iteration and generates an infinite sequence* $\{z^k = (t^k, x^k, y^k)\}$. *Moreover,* $0 < t^{k+1} \le t^k \le \bar{t}$ *and* $z^k \in \Omega$.

For any given $t \in R$, define $\psi_t(x, y) : R^n \times R^{md} \to R^+$ by

$$\psi_t(x, y) = \|G(z)\|^2. \tag{4.14}$$

It is easy to see that for any fixed $t \in R_{++}$, ψ_t is continuously differentiable with the gradient given by

$$\nabla \psi_t(x, y) = 2(G'_{(x,y)}(z))^T G(z), \tag{4.15}$$

where

$$G'_{(x,y)}(z) = \begin{pmatrix} -tI_n & A \\ P(z)A^T & I_{md} - P(z) \end{pmatrix}, \tag{4.16}$$

and $P(z)$ is defined in (4.5). By repeating the proof of Lemma 4.1, $G'_{(x,y)}(z)$ is nonsingular at any point $z = (t, x, y) \in R_{++} \times R^n \times R^{md}$. For any $z = (t, x, y) \in R \times R^n \times R^{md}$,

$$\psi(z) = t^2 + \psi_t(x, y). \tag{4.17}$$

It follows from Lemma 2.6 that we have the following.

LEMMA 4.3. *The set* $\mathcal{S} = \{(x, y) \in R^n \times R^{md} : \psi_0(x, y) = 0\}$ *is nonempty and bounded.*

LEMMA 4.4.

(i) *For any* $t > 0$ *and* $\alpha > 0$, *the level set*

$$L_t(\alpha) = \{(x, y) \in R^n \times R^{md} : \psi_t(x, y) \le \alpha\}$$

is bounded.

(ii) *For any* $0 < t_1 \le t_2$ *and* $\alpha > 0$, *the level set*

$$L_{[t_1,t_2]}(\alpha) = \{(x, y) \in R^n \times R^{md} : \psi_t(x, y) \le \alpha, t \in [t_1, t_2]\}$$

is bounded.

Proof. (i) For any $(x, y) \in L_t(\alpha)$,

$$\psi_t(x, y) = (Ay - tx)^2 + \sum_{i=1}^{m} \left(y_i - p(t, y_i + b_i - A_i^T x)\right)^2 \le \alpha.$$

So

$$\sum_{i=1}^{m} \left(y_i - p(t, y_i + b_i - A_i^T x)\right)^2 \le \alpha, \tag{4.18}$$

and

$$(Ay - tx)^2 \le \alpha. \tag{4.19}$$

From (4.18) y is bounded. It follows from (4.19) that x is bounded. Hence $L_t(\alpha)$ is bounded. Similarly, we can prove that (ii) holds. □

It follows from Lemma 4.4(i) that we have the following.

COROLLARY 4.5. *For any $t > 0$, $\psi_t(x, y)$ is coercive, i.e.,*

$$\lim_{\|(x,y)\| \to +\infty} \psi_t(x, y) = +\infty.$$

THEOREM 4.6.

(i) *An infinite sequence $\{z^k\} \subseteq R \times R^n \times R^{md}$ is generated by Algorithm* 4.1, *and*

$$\lim_{k \to +\infty} H(z^k) = 0 \text{ and } \lim_{k \to +\infty} t^k = 0. \tag{4.20}$$

Hence each accumulation point, say, $z^* = (0, x^*, y^*)$, of $\{z^k\}$ is a solution of $H(z) = 0$, and x^* and y^* are optimal solutions to problems (1.1) *and* (2.3), *respectively.*

(ii) *The sequence $\{z^k\}$ is bounded. Hence there exists at least an accumulation point, say, $z^* = (0, x^*, y^*)$, of $\{z^k\}$ such that x^* and y^* are optimal solutions to problems* (1.1) *and* (2.3), *respectively.*

(iii) *If problem* (1.1) *has a unique solution x^*, then*

$$\lim_{k \to +\infty} x^k = x^*.$$

Proof. The proof of (i) and (ii) is similar to that of Theorem 4.5 in [26], so we omit it. It is follows from (ii) that (iii) holds. □

Let $z^* = (0, x^*, y^*)$ and define

$$A(z^*) = \{\lim H'(t^k, x^k, y^k) : t^k \downarrow 0^+, x^k \to x^* \text{ and } y^k \to y^*\}. \tag{4.21}$$

Clearly, $A(z^*) \subseteq \partial H(z^*)$.

LEMMA 4.7. *If all $V \in A(z^*)$ are nonsingular, then there is a neighborhood $N(z^*)$ of z^* and a constant C such that for any $z = (t, x, y) \in N(z^*)$ with $t \ne 0$, $H'(z)$ is nonsingular and*

$$\|(H'(z))^{-1}\| \le C.$$

Proof. From Lemma 4.1, for any $z = (t, x, y) \in N(z^*)$ with $t \ne 0$, $H'(z)$ is nonsingular. If the conclusion is not true, then there is a sequence $\{z^k = (t^k, x^k, y^k)\}$ with all $t^k \ne 0$ such that $z^k \to z^*$, and $\|(H'(z^k))^{-1}\| \to +\infty$. Since H is locally Lipschitzian, ∂H is bounded in a neighborhood of z^*. By passing to a subsequence, we may assume that $H'(z^k) \to V$. Then V must be singular, a contradiction to the assumption of this lemma. This completes the proof. □

THEOREM 4.8. *Suppose that $z^* = (0, x^*, y^*)$ is an accumulation point of the infinite sequence $\{z^k\}$ generated by Algorithm 4.1 and all $V \in A(z^*)$ are nonsingular. Then the whole sequence $\{z^k\}$ converges to z^* quadratically.*

Proof. First, from Theorem 4.6 z^* is a solution of $H(z) = 0$. Then, from Lemma 4.7, for all z^k sufficiently close to z^*,

$$\|H'(z^k)^{-1}\| = O(1).$$

Because H is strongly semismooth at z^*, from Lemma 1.1, for z^k sufficiently close to z^*,

$$\begin{aligned} \|z^k + \Delta z^k - z^*\| &= \|z^k + H'(z^k)^{-1}[-H(z^k) + \beta_k \bar{z}] - z^*\| \\ &= O(\|H(z^k) - H(z^*) - H'(z^k)(z^k - z^*)\| + \beta_k \bar{t}\ ) \\ &= O(\|z^k - z^*\|^2) + O(\psi(z^k)), \end{aligned} \tag{4.22}$$

and H is locally Lipschitz continuous near z^*, i.e., for all z^k close to z^*,

$$\psi(z^k) = \|H(z^k)\|^2 = O(\|z^k - z^*\|^2). \tag{4.23}$$

Therefore, from (4.22) and (4.23), for all z^k sufficiently close to z^*,

$$\|z^k + \Delta z^k - z^*\| = O(\|z^k - z^*\|^2). \tag{4.24}$$

By following the proof of Theorem 3.1 in [27], for all z^k sufficiently close to z^*, we have

$$\|z^k - z^*\| = O(\|H(z^k) - H(z^*)\|). \tag{4.25}$$

Hence, for all z^k sufficiently close to z^*, we have

$$\begin{aligned} \psi(z^k + \Delta z^k) &= \|H(z^k + \Delta z^k)\|^2 \\ &= O(\|z^k + \Delta z^k - z^*\|^2) \\ &= O(\|z^k - z^*\|^4) \\ &= O(\|H(z^k) - H(z^*)\|^4) \\ &= O(\psi(z^k)^2). \end{aligned} \tag{4.26}$$

Therefore, for all z^k sufficiently close to z^* we have

$$z^{k+1} = z^k + \Delta z^k. \tag{4.27}$$

From (4.27) and (4.24),

$$\|z^{k+1} - z^*\| = O(\|z^k - z^*\|^2). \tag{4.28}$$

This completes the proof. ∎

Next, we study under what conditions all the matrices $V \in A(z^*)$ are nonsingular at a solution point $z^* = (0, x^*, y^*)$ of $H(z) = 0$.

PROPOSITION 4.9. *Suppose that $\|b_i - A_i^T x^*\| > 0$ for $i = 1, \dots, m$. Then all $V \in A(z^*)$ are nonsingular.*

Proof. Because $\|b_i - A_i^T x^*\| > 0$ for $i = 1, \dots, m$, $\|y_i^*\| = 1$ for $i = 1, \dots, m$. From (2.8) we have

$$\|y_i^* + b_i - A_i^T x^*\| > 1 \text{ for } i = 1, \dots, m.$$

Let $s_i = y_i + b_i - A_i^T x$ and $s_i^* = y_i^* + b_i - A_i^T x^*$ for $i = 1, \dots, m$. It is easy to see that for any $V \in A(z^*)$, there exists a sequence $\{z^k = (t^k, x^k, y^k)\}$ such that

$$V = \begin{pmatrix} 1 & 0 & 0 \\ -x^* & 0 & A \\ E^* & P^*A^T & I_{md} - P^* \end{pmatrix},$$

where

$$E^* = [E_1^*, \dots, E_m^*]^T,$$

$$(E_i^*)^T = \lim_{\substack{t^k \downarrow 0^+ \\ x^k \to x^* \\ y_i^k \to y_i^*}} \nabla \phi_t(t^k, s_i^k) \quad \text{for } i = 1, \dots, m,$$

and

$$P^* = \mathrm{Diag}\left(\frac{1}{\|s_i^*\|} I_d - \frac{1}{\|s_i^*\|^3} s_i^* (s_i^*)^T \right).$$

Let

$$M = \begin{pmatrix} 0 & A \\ P^*A^T & I_{md} - P^* \end{pmatrix}.$$

Hence, proving V is nonsingular is equivalent to proving M is nonsingular. Let

$$P_i^* = \frac{1}{\|s_i^*\|} I_d - \frac{1}{\|s_i^*\|^3} s_i^* (s_i^*)^T.$$

From Proposition 3.3, there exists a $d \times d$ matrix B_i^* such that

$$P_i^* = B_i \mathrm{Diag}(\lambda_j^i) B_i^T,$$

where $0 < \lambda_j^i < 1$ for $j = 1, \dots, d-1$ and $\lambda_d^i = 0$, and $B_i B_i^T = I_d$.

Let $B = \mathrm{Diag}(B_i)$ and $D = \mathrm{Diag}\left(\mathrm{Diag}(\lambda_j^i)\right)$. Then

$$M = \begin{pmatrix} I_n & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} 0 & AB \\ D(AB)^T & I_{md} - D \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & B^T \end{pmatrix}.$$

Let

$$N = \begin{pmatrix} 0 & AB \\ D(AB)^T & I_{md} - D \end{pmatrix}.$$

Then, proving M is nonsingular is equivalent to proving N is nonsingular.

Let $\bar{B} = \mathrm{Diag}(\bar{B}_i)$, where $\bar{B}_i$, $i = 1, \dots, m$, is a $d \times (d-1)$ matrix obtained by deleting the dth column of B_i, and $q = [q_1^T, q_2^T]^T = [q_1^T, q_{11}, \dots, q_{1d}, \dots, q_{m1}, \dots, q_{md}]^T \in R^n \times R^{md}$.

Let $Nq = 0$. Then we have

$$ABq_2 = 0, \tag{4.29}$$

and

$$D(AB)^T q_1 + (I_{md} - D)q_2 = 0. \tag{4.30}$$

Let

$$\bar{q}_2 = [q_{11}, \ldots, q_{1(d-1)}, \ldots, q_{m1}, \ldots, q_{m(d-1)}]^T \in R^{m(d-1)},$$

and

$$\bar{D} = \mathrm{Diag}\left(\mathrm{Diag}(\lambda_j^i, j = 1, \ldots, d-1)\right).$$

From (4.30) we have

$$q_{id} = 0 \quad \text{for } i = 1, \ldots, m, \tag{4.31}$$

and

$$D(A\bar{B})^T q_1 + (I_{m(d-1)} - \bar{D})\bar{q}_2 = 0. \tag{4.32}$$

Then, from (4.29) and (4.31),

$$A\bar{B}\bar{q}_2 = 0. \tag{4.33}$$

By following the proof of Lemma 4.1, we have $q_1 = 0$ and $\bar{q}_2 = 0$. Thus $q = 0$. This implies that N is nonsingular. So V is nonsingular. This completes the proof. ∎

PROPOSITION 4.10. *Let $M_0(z^*) = \{i : \|b_i - A_i^T x^*\| = 0, i = 1, \ldots, m\}$. If $\bar{A} = [A_i, i \in M_0(z^*)]$ is an $n \times n$ nonsingular matrix and $\|y_i^*\| < 1$ for $i \in M_0(z^*)$, then all $V \in A(z^*)$ are nonsingular.*

Proof. Without loss of generality, we suppose that $\|b_i - A_i^T x^*\| = 0$ for $i = 1, \ldots, j$ and $\|b_i - A_i^T x^*\| > 0$ for $i = j+1, \ldots, m$. Then $\|y_i^*\| < 1$ for $i = 1, \ldots, j$ and $\|y_i^*\| = 1$ for $i = j+1, \ldots, m$. From (2.8) we have

$$\|y_i^* + b_i - A_i^T x^*\| > 1, \text{ for } i = j+1, \ldots, m.$$

Let $s_i = y_i + b_i - A_i^T x$ and $s_i^* = y_i^* + b_i - A_i^T x^*$ for $i = 1, \ldots, m$. It is easy to see that for any $V \in A(z^*)$, there exists a sequence $\{z^k = (t^k, x^k, y^k)\}$ such that

$$V = \begin{pmatrix} 1 & 0 & 0 \\ -x^* & 0 & A \\ E^* & P^* A^T & I_{md} - P^* \end{pmatrix},$$

where

$$E^* = [E_1^*, \ldots, E_m^*]^T,$$

$$(E_i^*)^T = \lim_{\substack{t^k \downarrow 0^+ \\ x^k \to x^* \\ y_i^k \to y_i^*}} \nabla \phi_t(t^k, s_i^k) \quad \text{for } i = 1, \ldots, m,$$

and

$$P^* = \mathrm{Diag}\left(P_i^*\right),$$

$$P_i^* = I_d \text{ for } i = 1, \dots, j,$$

$$P_i^* = \frac{1}{\|s_i^*\|} I_d - \frac{1}{\|s_i^*\|^3} s_i^* (s_i^*)^T \text{ for } i = j+1, \dots, m.$$

Let

$$M = \begin{pmatrix} 0 & A \\ P^* A^T & I_{md} - P^* \end{pmatrix}.$$

Hence, proving V is nonsingular is equivalent to proving M is nonsingular.

Let

$$\tilde{A} = [A_{j+1}, \dots, A_m],$$

$$D = \text{Diag}\left(P_i^*, i = j+1, \dots, m\right),$$

and

$$q = [q_1^T, q_2^T, q_3^T]^T \in R^n \times R^n \times R^{md-n}.$$

Let $Mq = 0$. Then we have

$$\bar{A} q_2 + \tilde{A} q_3 = 0, \tag{4.34}$$

$$\bar{A}^T q_1 = 0, \tag{4.35}$$

and

$$D \tilde{A}^T q_1 + (I_{md-n} - D) q_3 = 0. \tag{4.36}$$

From (4.35) we have $q_1 = 0$. Then, from (4.36), $q_3 = 0$. It follows from (4.34) that $q_2 = 0$. Thus $q = 0$. This implies that M is nonsingular. So V is nonsingular. This completes the proof. □

By combining Theorem 4.8 and Propositions 4.9 and 4.10 we can directly obtain the following results.

THEOREM 4.11. *Suppose that $z^* = (0, x^*, y^*)$ is an accumulation point of the infinite sequence $\{z^k\}$ generated by Algorithm* 4.1. *If $\|b_i - A_i^T x^*\| > 0$ for $i = 1, \dots, m$, then the whole sequence $\{z^k\}$ converges to z^*, and the convergence is quadratic.*

THEOREM 4.12. *Suppose that $z^* = (0, x^*, y^*)$ is an accumulation point of the infinite sequence $\{z^k\}$ generated by Algorithm* 4.1. *Let $M_0(z^*) = \{i : \|b_i - A_i^T x^*\| = 0, i = 1, \dots, m\}$. If $\bar{A} = [A_i, i \in M_0(z^*)]$ is an $n \times n$ nonsingular matrix and $\|y_i^*\| < 1$ for $i \in M_0(z^*)$, then the whole sequence $\{z^k\}$ converges to z^* quadratically.*

5. Applications.

In this section, we will apply the algorithm proposed in section 4 to solve the ESFL problem, the EMFL problem, and the SMT problem under a given topology.

The ESFL problem. Let $a_1, a_2, \dots, a_m$ be m ($m \geq 2$) points in R^d, the d-dimensional Euclidean space. Let $\omega_1, \omega_2, \dots, \omega_m$ be m positive weights. Find a point $x \in R^d$ that minimizes

$$f(x) = \sum_{i=1}^m \omega_i \|x - a_i\|. \tag{5.1}$$

This is called the ESFL problem. For more information on this problem, see [23].

The ESFL problem can be easily transformed into a special case of problem (1.1) where $b_i = \omega_i a_i$ and $A_i^T = \omega_i I_d$, $i = 1, 2, \ldots, m$. Therefore, it follows from Theorems 4.6, 4.11, and 4.12 that we have the following theorem.

THEOREM 5.1. *For the ESFL problem, assume that an infinite sequence* $\{z^k\} \subseteq R \times R^d \times R^{md}$ *is generated by Algorithm* 4.1. *Then the following hold:*

(i) *There exists at least an accumulation* $z^* = (0, x^*, y^*)$ *such that* x^* *is an optimal solution to the ESFL problem.*

(ii) *Suppose* $\omega_i \|x^* - a_i\| > 0$ *for* $i = 1, \ldots, M$. *Then the whole sequence* $\{z^k\}$ *converges to* z^* *quadratically.*

(iii) *Suppose* $\omega_i \|x^* - a_i\| = 0$ *for some* i *and* $\omega_j \|x^* - a_j\| > 0$ *for all* $j \neq i$, *i.e., only the* i*th term is active, and* $\|y_i^*\| < 1$. *Then the whole sequence* $\{z^k\}$ *converges to* z^* *quadratically.*

The EMFL problem. Let $a_1, a_2, \ldots, a_M$ be M points in R^d, the d-dimensional Euclidean space. Let ω_{ji}, $j = 1, 2, \ldots, N$, $i = 1, 2, \ldots, M$, and υ_{jl}, $1 \leq j \leq l \leq N$, be given nonnegative numbers. Find a point $x = (x_1, x_2, \ldots, x_N) \in R^{dN}$ that minimizes

$$f(x) = \sum_{j=1}^{N} \sum_{i=1}^{M} \omega_{ji} \|x_j - a_i\| + \sum_{1 \leq j \leq l \leq N} \upsilon_{jl} \|x_j - x_l\|. \tag{5.2}$$

This is the so-called EMFL problem. For ease of notation, we assume that $\upsilon_{jj} = 0$ for $j = 1, 2, \ldots, N$ and $\upsilon_{jl} = \upsilon_{lj}$ for $1 \leq j \leq l \leq N$.

To transform the EMFL problem (5.2) into an instance of problem (1.1), we simply do the following. Let $x = (x_1, x_2, \ldots, x_N)$. It is clear that $x \in R^n$ where $n = dN$. For each nonzero ω_{ji}, there is a corresponding term of the Euclidean norm $\|c(\omega_{ji}) - A(\omega_{ji})^T x\|$ where $c(\omega_{ji}) = \omega_{ji} a_i$, and $A(\omega_{ji})^T$ is a row of N blocks of $d \times d$ matrices whose jth block is $\omega_{ji} I_d$ and whose other blocks are zero. For each nonzero υ_{jl}, there is a corresponding term of the Euclidean norm $\|c(\upsilon_{jl}) - A(\upsilon_{jl})^T u\|$ where $c(\upsilon_{jl}) = 0$ and $A(\upsilon_{jl})^T$ is a row of N blocks of $d \times d$ matrices whose jth and lth blocks are $-\upsilon_{jl} I_d$ and $\upsilon_{jl} I_d$, respectively, and whose other blocks are zero. Define the index set $\Sigma = \{1, 2, \ldots, \tau\}$, where the set $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_\tau\}$ is in one-to-one correspondence with the set of nonzero weights ω_{ji} and υ_{jl}, and then write problem (5.2) as follows.

Find a point $x = (x_1, x_2, \ldots, x_N) \in R^{dN}$ that minimizes

$$f(x) = \sum_{i=1}^{\tau} \|c_i - A_i^T x\|, \tag{5.3}$$

where for $i = 1, 2, \ldots, \tau$, $c_i \in R^d$, and $A_i \in R^{dN \times d}$. Therefore, it follows from Theorems 4.6, 4.11, and 4.12 that we have the following theorem.

THEOREM 5.2. *For the EMFL problem, assume that an infinite sequence* $\{z^k\} \subseteq R \times R^{dN} \times R^{\tau d}$ *is generated by Algorithm* 4.1. *Then the following hold:*

(i) *There exists at least an accumulation* $z^* = (0, x^*, y^*)$ *such that* x^* *is an optimal solution to the EMFL problem.*

(ii) *Suppose* $\|b_i - A_i^T x^*\| > 0$ *for* $i = 1, \ldots, \tau$. *Then the whole sequence* $\{z^k\}$ *converges to* z^* *quadratically.*

(iii) *Let* $\Sigma_0(x^*) = \{i \in \Sigma : \|b_i - A_i^T x^*\| = 0\}$. *Assume that* $|\Sigma_0(x^*)| = N$, *the matrices* A_i, $i \in \Sigma_0(x^*)$ *are linearly independent and* $\|y_i^*\| < 1$ *for* $i \in \Sigma_0(x^*)$. *Then the whole sequence* $\{z^k\}$ *converges to* z^* *quadratically.*

The SMT problem. The Euclidean SMT problem is given by a set of points $P = \{p_1, p_2, \ldots, p_N\}$ in the Euclidean plane and asks for the shortest planar straight-line graph spanning P. The solution takes the form of a tree, called the SMT, that includes all the given points, called *regular points*, along with some extra vertices, called *Steiner points*. It is known that there are at most $N-2$ Steiner points and the degree of each Steiner point is at most 3; see [17]. A full Steiner topology of point set P is a tree graph whose vertex set contains P and $N-2$ Steiner points and where the degree of each vertex in P is exactly 1 and the degree of each Steiner vertex is exactly 3.

Computing an SMT for a given set of N points in the Euclidean plane is NP-hard. However, the problem of computing the shortest network under a given full Steiner topology can be solved efficiently. We can transform this problem into the following problem; see [35] for more detail.

Find a point $x = (x_1, x_2, \ldots, x_{N-2}) \in R^{2N-4}$ that minimizes

$$f(x) = \sum_{i=1}^{m} \|c_i - A_i^T x\|, \tag{5.4}$$

where for $i = 1, 2, \ldots, m$, $c_i \in R^2$, and $A_i \in R^{2(N-2)\times 2}$. Therefore, it follows from Theorems 4.6, 4.11, and 4.12 that we have the following theorem.

THEOREM 5.3. *For the problem of computing the shortest network under a given full Steiner topology, assume that an infinite sequence $\{z^k\} \subseteq R \times R^{2N-4} \times R^{4N-6}$ is generated by Algorithm* 4.1. *Then the following hold:*

(i) *There exists at least an accumulation $z^* = (0, x^*, y^*)$ such that x^* is an optimal solution to the EMFL problem.*

(ii) *Suppose $\|c_i - A_i^T x^*\| > 0$ for $i = 1, \ldots, m$. Then the whole sequence $\{z^k\}$ converges to z^* quadratically.*

(iii) *Let $M_0(x^*) = \{i : \|b_i - A_i^T x^*\| = 0, i = 1, 2, \ldots, m\}$. Assume that $|M_0(x^*)| = N$, the matrices A_i, $i \in M_0(x^*)$, are linearly independent and $\|y_i^*\| < 1$ for $i \in M_0(x^*)$. Then the whole sequence $\{z^k\}$ converges to z^* quadratically.*

6. Numerical experiments.

Algorithm 4.1 was implemented in MATLAB and was run on a DEC Alpha Server 8200 for the following examples, where Examples 1(a)–5 and 8 are taken from [25] and Examples 6 and 7 from [35]. Throughout the computational experiments, unless otherwise stated, we used the following parameters:

$$\delta = 0.5,\ \sigma = 0.0005,\ \bar{t} = 0.002,\ y^0 = 0, \quad \text{and} \quad \gamma = 0.5.$$

We terminated our iteration when one of the following conditions was satisfied:

(1) $k > 50$;

(2) $\text{relgap}(x^k, y^k) \le$ 1e-8, $\|Ay\| \le$ 1e-12, and $\max_{1\le i\le m} \|y_i\| \le$ 1+1e-8;

(3) $ls > 20$,

where *ls* was the number of line search at each step and

$$\text{relgap}(x, y) = \frac{\left|\sum_{i=1}^m \|b_i - A_i^T x\| - b^T y\right|}{\sum_{i=1}^m \|b_i - A_i^T x\| + 1}.$$

The numerical results which we obtained are summarized in Table 1. In this table, n, d, and m specify the problem dimensions, Iter denotes the number of iterations, which is also equal to the number of Jacobian evaluations for the function

TABLE 1
Numerical results for Algorithm 4.1.

Example	n	d	m	Iter	NH	N0	$f(x^k)$	relgap	$\|Ay\|$	$\max_{1\le i\le m} \|y_i\|$
1(a)	2	2	3	7	12	1	2.828427	0	1.11e-16	1.00
1(b)	2	2	3	6	12	1	2.828427	4.60e-12	1.19e-13	1.00
1(c)	2	2	3	6	12	1	2.828427	4.60e-12	1.19e-13	1.00
1(d)	2	2	3	6	12	1	2.828427	4.58e-12	1.18e-13	1.00
2	2	2	3	7	14	0	2.732051	0	0	1.00
3	2	2	3	7	12	0	2.828427	1.16e-16	0	1.00
4	2	2	3	7	12	1	2.828427	1.74e-15	1.11e-16	1.00
5	10	2	55	12	27	2	226.2084	2.84e-14	6.26e-13	1.00
6	16	2	17	9	20	4	25.35607	5.80e-15	2.40e-15	1.00
7	4	2	5	4	5	1	400.0200	3.83e-15	6.75e-15	1.00
8	3	3	100	11	44	0	558.6450	8.13e-16	3.83e-14	1.00

TABLE 2
Output of Algorithm 4.1 *for Example* 5.

k	relgap	$\|Ay^k\|$	$\max_{1\le i\le m} \|y_i^k\|$	t^k	N0	δ^{j_k}
1	4.91e-01	1.47e-03	3.35e+00	1.50e-03	0	5.0e-01
2	4.72e-01	1.72e-03	3.28e+00	1.48e-03	0	3.1e-02
3	4.70e-01	1.75e-03	3.27e+00	1.48e-03	0	3.9e-03
4	1.04e-01	1.08e-02	2.84e+00	1.00e-03	0	1.0e+00
5	1.08e-03	1.07e-02	3.80e+00	1.00e-03	0	1.0e+00
6	4.27e-03	9.21e-03	1.56e+00	1.00e-03	0	1.0e+00
7	4.00e-04	3.74e-03	1.10e+00	4.07e-04	0	1.0e+00
8	7.82e-05	3.20e-04	1.03e+00	3.44e-05	0	1.0e+00
9	4.40e-06	7.91e-06	1.02e+00	9.00e-07	0	1.0e+00
10	1.66e-07	3.79e-06	1.00e+00	4.13e-07	2	1.0e+00
11	1.08e-09	1.30e-10	1.00e+00	1.44e-11	2	1.0e+00
12	2.84e-14	6.26e-13	1.00e+00	6.82e-14	2	1.0e+00

H, NH denotes the number of function evaluations for the function H, N0 indicates the number of norms that are zero at the optimal solution, more precisely, which is interpreted as being zero if it is less than the tolerance 10^{-10}, $f(x^k)$ denotes the value of $f(x)$ at the final iteration, and relgap denotes the relative duality gap. The results reported in Table 1 show that this method is extremely promising. The algorithm was able to solve all examples in less than 15 iterations. Tables 2 and 3 give more detailed results for Examples 5 and 6, which show the quadratic convergence of this method. For Examples 6 and 7, the number of iterations required by our algorithm is fewer than that required by the algorithm proposed in [35].

The first few examples are of the following form:

$$\begin{aligned} &n = 2, \quad d = 2, \quad m = 3, \\ &A_1 = I, \quad A_2 = \omega I, \quad A_3 = I, \\ &b_1 = [-1, 0]^T, \quad b_2 = [0, \omega]^T, \quad b_3 = [1, 0]^T. \end{aligned} \tag{6.1}$$

Example 1(a). This is given by (6.1) with $\omega = 2$ and solution $x^* = [0.0, 1.0]^T$. The starting point $x^0 = [3.0, 2.0]^T$.

Example 1(b). Same as Example 1(a), except $x^0 = [1.0, 1.0 \times 10^{-6}]^T$.

Example 1(c). Same as Example 1(a), except $x^0 = [1.000001, -1.0 \times 10^{-6}]^T$.

Example 1(d). Same as Example 1(a), except $x^0 = [1.001, -1.0 \times 10^{-3}]^T$.

Example 2. This is given by (6.1) with $\omega = 1$ and solution $x^* = [0.0, 0.577350]^T$. The starting point $x^0 = [3.0, 2.0]^T$.

TABLE 3
Output of Algorithm 4.1 for Example 6.

k	relgap	$\|Ay^k\|$	$\max_{1\le i\le m} \|y_i^k\|$	t^k	N0	δ^{j_k}
1	7.52e-01	1.97e-03	1.65e+00	1.75e-03	0	2.5e-01
2	5.69e-01	1.45e-02	1.62e+00	1.66e-03	0	1.2e-01
3	2.25e-01	2.60e-02	1.44e+00	1.49e-03	0	2.5e-01
4	1.77e-01	2.37e-02	1.42e+00	1.43e-03	0	1.2e-01
5	4.39e-02	1.70e-02	1.17e+00	1.00e-03	0	1.0e+00
6	5.22e-03	4.05e-03	1.03e+00	2.26e-04	0	1.0e+00
7	1.01e-04	6.83e-05	1.00e+00	3.56e-06	0	1.0e+00
8	4.10e-08	1.62e-08	1.00e+00	8.36e-10	2	1.0e+00
9	5.80e-15	2.40e-15	1.00e+00	1.34e-16	4	1.0e+00

TABLE 4
Weights: New to new and new to existing.

	New					Existing								
New	1	2	3	4	5	1	2	3	4	5	6	7	8	9
1		1	1	1	1	2	2	1	1	1	1	1	1	1
2			1	10^{-2}	10^{-1}	1	1	2	2	1	1	1	1	1
3				10^{-2}	10^{-1}	1	1	1	1	2	2	1	1	1
4					10^{-1}	1	1	1	1	1	1	2	2	1
5						1	1	1	1	1	1	1	1	2

TABLE 5
Existing facility locations.

	1	2	3	4	5	6	7	8	9
Component 1	0	2	6	6	8	7	0	0	0
Component 2	0	4	2	10	8	7	1	2	3

Example 3. This is given by (6.1) with $\omega = 1.414$ and solution $x^* = [0.0, 0.999698]^T$. The starting point $x^0 = [3.0, 2.0]^T$.

Example 4. This is given by (6.1) with $\omega = 1.415$ and solution $x^* = [0.0, 1.0]^T$. The starting point $x^0 = [3.0, 2.0]^T$.

Example 5. This is a multifacility location problem. The objective is to choose five new facilities in the plane (i.e., vectors in R^2) to minimize a weighted sum of distances between each pair of new facilities plus a weighted sum of distances between each of the new facilities and each of the existing facilities (i.e., given vectors in R^2). Tables 4 and 5 complete the description of the problem. The solution is

$$x^* = [2.03865, 3.65117; 2.24659, 3.75886; 2.24659, 3.75886; 1.45825, 2.96083; 2.03865, 3.65117]^T.$$

The starting point $x^0 = [1,1;1,1;1,1;1,1;1,1]^T$.

Example 6. This is an SMT problem. This example contains 10 regular points. The coordinates of the 10 regular points are given in Table 6. The tree topology is given in Table 6 where for each edge, indices of its two vertices are shown next to the index of the edge. This topology is the best topology obtained by a branch-and-bound algorithm. Therefore, the shortest network under this topology is actually the SMT problem for the given 10 regular points. The starting point $x^0 = [1,1;1,1;1,1;1,1;1,1;1,1;1,1;1,1]^T$.

Example 7. This is an SMT problem. This example contains four regular points. The coordinates of the four regular points and the tree topology are given in Table 7.

TABLE 6
The topology and the coordinates of the ten regular point in Example 6.

Point-index	x-coord	y-coord	Point-index	x-coord	y-coord
9	2.309469	9.208211	14	7.598152	0.615836
10	0.577367	6.480938	15	8.568129	3.079179
11	0.808314	3.519062	16	4.757506	3.753666
12	1.685912	1.231672	17	3.926097	7.008798
13	4.110855	0.821114	18	7.436490	7.683284
Edge-index	ea-index	eb-index	Edge-index	ea-index	eb-index
1	9	7	10	18	8
2	10	1	11	5	6
3	11	2	12	6	4
4	12	3	13	4	3
5	13	4	14	3	2
6	14	5	15	2	1
7	15	5	16	1	7
8	16	6	17	7	8
9	17	8			

TABLE 7
The topology and the coordinates of the four regular point in Example 7.

Point-index	x-coord	y-coord	Point-index	x-coord	y-coord
3	−100.0	1.0	5	−100.0	−1.0
4	100.0	1.0	6	100.0	−1.0
Edge-index	ea-index	eb-index	Edge-index	ea-index	eb-index
1	3	1	4	6	2
2	4	1	5	1	2
3	5	2			

The starting point $x^0 = [1, 1; 1, 1]^T$.

Example 8.

$$n = 3, \quad d = 3, \quad m = 100.$$

$$A_i = I, \ i = 1, 2, \ldots, m, \text{ except } A_i = 100I \text{ if } i \bmod 10 = 1.$$

The elements of b_i, $i = 1, 2, \ldots, m$, are generated randomly. We use the following pseudorandom sequence:

$$\psi_0 = 7, \psi_{i+1} = (445\psi_i + 1) \bmod \quad 4096, i = 1, 2, \ldots,$$

$$\bar{\psi}_i = \frac{\psi_i}{4096}, \qquad i = 1, 2, \ldots.$$

The elements of b_i, $i = 1, 2, \ldots, m$, are successively set to be $\bar{\psi}_1, \bar{\psi}_2, \ldots$ in the order $(b_1)_1, \ldots, (b_1)_d, (b_2)_1, \ldots, (b_m)_d$, except that the appropriate random number is multiplied by 100 to given $(b_i)_j$ if $i \bmod 10 = 1$.

The solution $x^* = [0.586845, 0.480333, 0.509340]^T$. The initial point x^0 is set to b_m.

7. Conclusions. In this paper we presented a smoothing Newton method for the problem of minimizing a sum of Euclidean norms by applying the smoothing Newton method proposed by Qi, Sun, and Zhou [28] directly to a system of strongly semismooth equations derived from primal and dual feasibility and a complementarity condition, and proved that this method was globally and quadratically convergent. It is deserved to point out that in our method we can control the smoothing parameter t in such a way that it converges to zero neither too quickly nor too slowly by using a particularly designed Newton equation and a line search model; see (4.11) and (4.12). Numerical results indicated that our algorithm was extremely promising. It will be an interesting work to compare this method with some existing methods, e.g., the primal-dual interior-point method proposed in [3]. However, we have been unable to do this because no code is available.

Consider the problem of minimizing a sum of Euclidean norms subject to linear equality constraints:

$$\min \left\{ \sum_{i=1}^{m} \|b_i - A_i^T x\|, \quad E^T x = b_e,\ x \in R^n \right\}, \tag{7.1}$$

where $E \in R^{n\times d}$ is an $n \times d$ matrix with full column rank and $b_e \in R^d$. In [2], Andersen and Christiansen transformed the problem (7.1) to the problem (1.1) based on the l_1 penalty function approach. So we can also apply the algorithm proposed in section 4 to solve problem (7.1).

Acknowledgments. The authors would like to thank Steve Wright, the two referees, Defeng Sun, Houduo Qi, and Song Xu for their helpful comments. Thanks also go to K. D. Andersen and E. Christiansen for providing the authors with some references.

REFERENCES

[1] K. D. Andersen, *An efficient Newton barrier method for minimizing a sum of Euclidean norms*, SIAM J. Optim., 6 (1996), pp. 74–95.

[2] K. D. Andersen and E. Christiansen, *Minimizing a sum of norms subject to linear equality constraints*, Comput. Optim. Appl., 11 (1998), pp. 65–79.

[3] K. D. Andersen, E. Christiansen, A. R. Conn, and M. L. Overton, *An efficient primal-dual interior-point method for minimizing a sum of Euclidean norms*, SIAM J. Sci. Comput., 22 (2000), pp. 243–262.

[4] P. H. Calamai and A. R. Conn, *A stable algorithm for solving the multifacility location problem involving Euclidean distances*, SIAM J. Sci. Statist. Comput., 1 (1980), pp. 512–526.

[5] P. H. Calamai and A. R. Conn, *A projected Newton method for l_p norm location problems*, Math. Programming, 38 (1987), pp. 75–109.

[6] R. Chandrasekaran and A. Tamir, *Open questions concerning Weiszfeld's algorithm for the Fermat-Weber location problem*, Math. Programming, 44 (1989), pp. 293–295.

[7] R. Chandrasekaran and A. Tamir, *Algebraic optimization: The Fermat-Weber location problem*, Math. Programming, 46 (1990), pp. 219–224.

[8] B. Chen and X. Chen, *A global and local superlinear continuation-smoothing method for $P_0 + R_0$ NCP or monotone NCP*, SIAM J. Optim., 9 (1999), pp. 624–645.

[9] C. Chen and O. L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comput. Optim. Appl., 5 (1996), pp. 97–138.

[10] X. Chen, L. Qi, and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Math. Comp., 67 (1998), pp. 519–540.

[11] X. Chen and Y. Ye, *On homotopy-smoothing methods for box-constrained variational inequalities*, SIAM J. Control Optim., 37 (1999), pp. 589–616.

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

[13] J. W. Eyster, J. A. White, and W. W. Wierwille, *On solving multifacility location problems using a hyperboloid approximation procedure*, AIIE Trans., 5 (1973), pp. 1–6.

[14] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Programming, 75 (1996), pp. 407–439.

[15] F. Facchinei and C. Kanzow, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Math. Programming, 76 (1997), pp. 493–512.

[16] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Programming, 76 (1997), pp. 513–532.

[17] E. N. Gilbert and H. O. Pollak, *Steiner minimal trees*, SIAM J. Appl. Math., 16 (1968), pp. 1–29.

[18] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 707–716.

[19] H. Jiang, *Smoothed Fischer-Burmeister Equation Methods for the Complementarity Problem*, Department of Mathematics, The University of Melbourne, Victoria, Australia, Preprint, 1997.

[20] K. Hotta and A. Yoshise, *Global convergence of a class of non-interior-point algorithms using Chen-Harker-Kanzow functions for nonlinear complementarity problems*, Math. Programming, 86 (1999), pp. 105–133.

[21] H. Jiang and L. Qi, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control Optim., 35 (1997), pp. 178–193.

[22] H. W. Kuhn, *A note on Fermat's problem*, Math. Programming, 4 (1973), pp. 98–107.

[23] R. F. Love, J. G. Morris, and G. O. Wesolowsky, *Facility Location: Models and Methods*, North–Holland, Amsterdam, 1988.

[24] L. M. M. Ostresh, *The multifacility location problem: Applications and decent theorems*, Journal of Regional Science, 17 (1977), pp. 409–419.

[25] M. L. Overton, *A quadratically convergent method for minimizing a sum of Euclidean norms*, Math. Programming, 27 (1983), pp. 34–63.

[26] H.-D. Qi, *A regularized smoothing Newton method for box constrained variational inequality problems with P_0-functions*, SIAM J. Optim., 10 (2000), pp. 315–330.

[27] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

[28] L. Qi, D. Sun, and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Programming, 87 (2000), pp. 1–35.

[29] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.

[30] J. B. Rosen and G. L. Xue, *On the convergence of Miehle's algorithm for the Euclidean multifacility location problem*, Oper. Res., 40 (1992), pp. 188–191.

[31] J. B. Rosen and G. L. Xue, *On the convergence of a hyperboloid approximation procedure for solving the perturbed Euclidean multifacility location problem*, Oper. Res., 41 (1993), pp. 1164–1171.

[32] D. Sun, *A regularization Newton method for solving nonlinear complementarity problems*, Appl. Math. Optim., 40 (1999), pp. 315–339.

[33] C. Y. Wang et al, *On the convergence and rate of convergence of an iterative algorithm for the plant location problem*, Qufu Shifan Xuebao, 2 (1975), pp. 14–25 (in Chinese).

[34] E. Weiszfeld, *Sur le point par lequel la somme des distances de n points donnes est minimum*, Tohoku Math. J., 43 (1937), pp. 355–386.

[35] G. Xue and Y. Ye, *An efficient algorithm for minimizing a sum of Euclidean norms with applications*, SIAM J. Optim., 7 (1997), pp. 1017–1036.

[36] G. Xue and Y. Ye, *An efficient algorithm for minimizing a sum of p-norms*, SIAM J. Optim., 10 (2000), pp. 551–579.

[37] N. Yamashita and M. Fukushima, *Modified Newton methods for solving semismooth reformulations of monotone complementarity problems*, Math. Programming, 76 (1997), pp. 273–284.

[38] G. Zhou, D. Sun, and L. Qi, *Numerical experiments for a class of squared smoothing Newton methods for box constrained variational inequality problems*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Appl. Optim. 22, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 421–441.