

## AN EFFICIENT ALGORITHM FOR MINIMIZING A SUM OF $p$ -NORMS\*

GUOLIANG XUE<sup>†</sup> AND YINYU YE<sup>‡</sup>

**Abstract.** We study the problem of minimizing a sum of  $p$ -norms where  $p$  is a fixed real number in the interval  $[1, \infty]$ . Several practical algorithms have been proposed to solve this problem. However, none of them has a known polynomial time complexity. In this paper, we transform the problem into standard conic form. Unlike those in most convex optimization problems, the cone for the  $p$ -norm problem is *not* self-dual unless  $p = 2$ . Nevertheless, we are able to construct two logarithmically homogeneous self-concordant barrier functions for this problem. The barrier parameter of the first barrier function does not depend on  $p$ . The barrier parameter of the second barrier function increases with  $p$ . Using both barrier functions, we present a primal-dual potential reduction algorithm to compute an  $\epsilon$ -optimal solution in polynomial time that is independent of  $p$ . Computational experiences with a Matlab implementation are also reported.

**Key words.** shortest network under a given topology, facilities location, Steiner minimum trees, minimizing a sum of norms, primal-dual potential reduction algorithms, polynomial time algorithms

**AMS subject classifications.** 68Q20, 68Q25, 90C25, 90C35

**PII.** S1052623497327088

**1. Introduction.** Let  $c_1, c_2, \dots, c_m \in R^d$  be column vectors in the Euclidean  $d$ -space and  $A_1, A_2, \dots, A_m \in R^{n \times d}$  be  $n$ -by- $d$  matrices each having full column rank. We want to find a point  $u \in R^n$  such that the following sum of  $p$ -norms,  $p \geq 1$ , is minimized:

$$(1.1) \quad \begin{aligned} \min \quad & \sum_{i=1}^m \|c_i - A_i^T u\|_p \\ \text{s.t.} \quad & u \in R^n. \end{aligned}$$

We recall that  $\|\bullet\|_p$  is the Hölder or  $p$ -norm [14] defined by

$$(1.2) \quad \|x\|_p = (|x_1|^p + \dots + |x_n|^p)^{1/p}.$$

It is clear that  $u = 0$  is an optimal solution to (1.1) when all of the  $c_i$  are zero. Therefore, we will assume in the rest of this paper that not all of the  $c_i$  are zero. Problem (1.1) is a convex programming problem. Thus, it can be generally solved in “polynomial time.” In this paper, we present a conic formulation of the problem, develop a specialized interior point algorithm, and analyze its computational complexity, exploring the special tree structure of the problem. One interesting feature is that the  $p$ -order cone involved in our formulation, unlike those in almost all current problems solved by interior point algorithms, is *not* self-dual. This presents certain difficulties in developing and analyzing the algorithm.

---

\*Received by the editors September 5, 1997; accepted for publication (in revised form) July 27, 1998; published electronically February 24, 2000.

<http://www.siam.org/journals/siopt/10-2/32708.html>

<sup>†</sup>Department of Computer Science, The University of Vermont, Burlington, VT 05405-0156 (xue@cs.uvm.edu). The research of this author was supported in part by U.S. Army Research Office grant DOD DAAH04-96-1-0233 and by National Science Foundation grants NSF ASC-9409285 and NSF OSR-9350540.

<sup>‡</sup>Department of Management Sciences, The University of Iowa, Iowa City, IA 52242-1000 (yyye@dollar.biz.uiowa.edu). The research of this author was supported in part by NSF grants DMI-9522507, DMS-9703490, and OSR-9350540.

Problem (1.1) has a long history, dating back to the 17th century, when Fermat [19] studied the problem of finding the shortest network interconnecting three points on the Euclidean plane, where a fourth point may be added to minimize the sum of lengths of the network. The Fermat problem was generalized to the Euclidean facilities location problem and the Euclidean Steiner minimal tree problem. Those problems were in turn further generalized to cases where distances are measured using weighted  $p$ -norms.

The facilities location problem is one of locating  $N$  new facilities with respect to  $M$  existing facilities, the locations of which are known. The problem consists of finding locations of new facilities that will minimize the sum of (nonnegatively) weighted distances between the new and existing facilities, and between the new facilities. If there is only one new facility ( $N = 1$ ), the problem is called a single facility location (SFL) problem. If there is more than one new facility ( $N \geq 2$ ), the problem is called a multifacility location (MFL) problem.

For the general SFL problem with Euclidean norm, Weiszfeld [33] gave a simple closed-form iterative algorithm in 1937. This work started a chain of research on this topic [19, 27, 17, 32, 5, 6, 20]. Miehle [23] was the first to propose an extension of the Weiszfeld algorithm for the SFL problem to solve MFL problems. Again, a number of important results were obtained along this line [27, 29, 34, 10, 30]. For more details on location problems, see the books by Francis, McGinnis, and White [11] and Love, Morris, and Wesolowsky [21].

Practical algorithms for solving these problems began with the work of Calamai and Conn [3, 4] and Overton [28], where they proposed projected Newton algorithms with quadratic rate of convergence. They also generalized the location problems to one of minimizing a sum of norms. In recent years, several complexity results and numerically stable algorithms have been obtained for problem (1.1) with  $p = 2$  using techniques of interior point algorithms. In [35], Xue, Rosen, and Pardalos showed that the dual of the Euclidean MFL problem is the minimization of a linear function subject to linear and convex quadratic constraints and that therefore it can be solved using interior point techniques in polynomial time. More recently, Andersen [1] used the HAP idea [10] to smooth the objective function by introducing a perturbation  $\epsilon > 0$  and applied a Newton barrier method to solve the problem. Andersen and Christiansen [2] and Conn and Overton [7] also proposed a primal-dual method based on the  $\epsilon$ -perturbation and presented impressive computational results.

For  $p \geq 1$ , den Hertog et al. [8, 9] (also see references therein) presented a polynomial time interior point Newton barrier method for solving a closely related problem, where the objective function is

$$\sum_{i=1}^m \|c_i - A_i^T u\|_p^p,$$

that is, the sum of the  $p$ -powers of  $p$ -norms. This objective function is equivalent to ours only if  $m = 1$ . Nesterov and Nemirovskii [25, 24] also addressed the problem with  $m = 1$ . Their constructed barrier function, although different from the one used in our conic formulation, influenced our study of this problem.

In a recent paper [36], we studied the problem of minimizing a sum of Euclidean norms. The problem was formulated in standard conic form and a polynomial time primal-dual potential reduction algorithm was presented. A nice property of the Euclidean case is that the second-order cone is self-dual. We have taken advantage of this fact in our work in [36], based on the self-dual theory and primal-dual algorithm

developed by Nesterov and Todd [26] (see also Kojima, Shindoh, and Hara [18]).

In some applications, the problems are better modeled with  $p$ -norms where  $p \neq 2$ . For example, in VLSI design, the 1-norm or Manhattan distance is used. In transportation, several general  $p$ -norms are used [21]. Since the  $p$ -order cone of the conic form of (1.1) is *not* self-dual unless  $p = 2$ , a natural question to ask is the following: *Can the techniques used for minimizing a sum of Euclidean norms be generalized to minimize a sum of  $p$ -norms in polynomial time?* The goal of this paper is to answer the above question affirmatively. Specifically, we present a primal-dual potential reduction algorithm that computes an  $\epsilon$ -optimal solution in at most

$$O(\sqrt{md}(\log(\bar{c}/\epsilon) + \log(md)))$$

iterations, where  $\bar{c} = \max_{1 \leq i \leq m} \|c_i\|$ . Note that this bound is independent of  $p$  and is increased only by a factor  $\sqrt{d}$  compared to the bound for  $p = 2$  [36].

The rest of this paper is organized as follows. In section 2, the basic problem (1.1) is transformed into a standard convex programming problem in conic form. In section 3, we develop two logarithmically homogeneous self-concordant barrier functions for this problem. In section 4, we present a primal-dual potential reduction algorithm for solving the problem. In section 5, we discuss the computational complexity and simplifications of the potential reduction algorithm. In section 6, we present applications to the SFL problem, the MFL problem, and the Steiner minimal tree (SMT) problem. In section 7, we present some computational examples on SMT problems. We conclude this paper in section 8.

**2. Conic formulation.** We will call problem (1.1) the *basic problem* in the rest of our paper. This problem can be formulated as the maximization of a linear function subject to affine and convex cone constraints as follows.

$$(2.1) \quad \begin{aligned} \max \quad & - \sum_{i=1}^m t_i \\ \text{s.t.} \quad & t_1 \geq \|c_1 - A_1^T u\|_p, \\ & t_2 \geq \|c_2 - A_2^T u\|_p, \\ & \vdots \\ & t_m \geq \|c_m - A_m^T u\|_p, \end{aligned}$$

where  $t_i \in R, i = 1, 2, \dots, m$ .

In the rest of this paper, when we represent a large matrix with several small matrices, we will use a semicolon to represent column concatenation and a comma to represent row concatenation. This notation also applies to vectors. We will use  $0$  to represent a column vector whose elements are all zero, and  $e$  the vector of all ones. We will use  $R_+$  to represent the set of nonnegative real numbers.

In this section, we will transform our basic problem (1.1) into a standard convex programming problem in conic form, where the cone and its associated barrier are *not self-dual* unless  $p = 2$ . (It is worthwhile to mention that, for  $p = 1$  or  $p = \infty$ , the problem can be reformulated as a linear program or second-order conic program of  $md$  variables; thus, the reformulation becomes self-dual and the symmetric-scaling algorithm applies.)

Consider the  $p$ -order cone, where  $p \geq 1$ ,

$$K = \left\{ (t \in R_+, s \in R^d) : t^p \geq \sum_{j=1}^d |s_j|^p \right\}.$$

Its interior is

$$\text{int}K := \{(t \in R_+, s \in R^d) : t > \|s\|_p\}.$$

The dual of  $K$  is

$$K^* = \left\{ (\tau \in R_+, x \in R^d) : \tau^q \geq \sum_{j=1}^d |x_j|^q \right\},$$

where  $\frac{1}{p} + \frac{1}{q} = 1$ .

Now let

$$\mathcal{B} = \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \\ 0_n \end{pmatrix} \in R^{m+n}, \quad \mathcal{C} = \begin{pmatrix} (0; c_1) \\ (0; c_2) \\ \vdots \\ (0; c_m) \end{pmatrix} \in R^{m+md},$$

and

$$\mathcal{A}^T = \begin{pmatrix} -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & A_1^T \\ 0 & -1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & A_2^T \\ & & \ddots & & \\ 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & \cdots & 0 & A_m^T \end{pmatrix} \in R^{(m+md) \times (m+n)}.$$

Then, problem (1.1) or (2.1) can be written in the standard (dual) form

$$(2.2) \quad \begin{aligned} & \max \quad \mathcal{B}^T(t_1; t_2; \dots; t_m; u) \\ & \text{s.t.} \quad \begin{pmatrix} (t_1; s_1) \\ (t_2; s_2) \\ \vdots \\ (t_m; s_m) \end{pmatrix} = \mathcal{C} - \mathcal{A}^T(t_1; t_2; \dots; t_m; u), \\ & \quad (t_i; s_i) \in K, \quad i = 1, 2, \dots, m, \end{aligned}$$

where  $u \in R^n$  and  $t_i \in R_+, s_i \in R^d, i = 1, 2, \dots, m$ .

Let  $(\tau_1; x_1), (\tau_2; x_2), \dots, (\tau_m; x_m) \in R^{d+1}$ . Then its corresponding primal problem is

$$(2.3) \quad \begin{aligned} & \min \quad \mathcal{C}^T((\tau_1; x_1); (\tau_2; x_2); \dots; (\tau_m; x_m)) \\ & \text{s.t.} \quad \mathcal{A}((\tau_1; x_1); (\tau_2; x_2); \dots; (\tau_m; x_m)) = \mathcal{B}, \\ & \quad (\tau_i; x_i) \in K^*, \quad i = 1, 2, \dots, m. \end{aligned}$$

Thus, using  $\mathcal{S} := ((t_1; s_1); (t_2; s_2); \dots; (t_m; s_m)), \mathcal{Y} := (t_1; t_2; \dots; t_m; u), \mathcal{X} := ((\tau_1; x_1); (\tau_2; x_2); \dots; (\tau_m; x_m)), \mathcal{K} := K^m := K \times K \times \dots \times K$ , and  $\mathcal{K}^* := (K^*)^m := K^* \times K^* \times \dots \times K^*$ , we can write the two problems (2.3) and (2.2) as

$$(P) \quad \begin{aligned} & \min \quad \mathcal{C}^T \mathcal{X} \\ & \text{s.t.} \quad \mathcal{A} \mathcal{X} = \mathcal{B}, \\ & \quad \mathcal{X} \in \mathcal{K}^*, \end{aligned}$$

and

$$(D) \quad \begin{aligned} & \max \mathcal{B}^T \mathcal{Y} \\ & \text{s.t. } \mathcal{S} = \mathcal{C} - \mathcal{A}^T \mathcal{Y}, \\ & \mathcal{S} \in \mathcal{K}. \end{aligned}$$

This is the pair of problems (P) and (D) in Nesterov and Nemirovskii [25]. Unlike those in most optimization problems,  $\mathcal{K} \neq \mathcal{K}^*$  (unless  $p = 2$ ) here. Therefore, the symmetric primal-dual techniques, developed by, e.g., Nesterov and Todd [26], are no longer applicable. However, we can still use interior point algorithms to compute an  $\epsilon$ -optimal solution of the problem in polynomial time.

**3. Barrier functions for  $p$ -order cones.** The key to solving problems (P) and (D) is to construct a simple and efficient barrier function for  $K$ , when  $p \neq 2$ . This has been an open question. In this section we present two barrier functions and analyze their barrier parameters. To simplify notation, we will use  $s$  and  $z$  to denote vectors in  $R^d$  and use  $s_j$  and  $z_j$  to denote the  $j$ th single component of the vectors  $s$  and  $z$ . This notation is different than that used in the previous sections (where  $s_i$  stood for a vector in  $R^d$ ) and will be used *only in this section*.

**3.1. Barrier function I.** One barrier function is constructed from the following convex set:

$$G_p = \left\{ (s \in R^d, z \in R^d) : z_i \geq |s_i|^p, i = 1, \dots, d, \sum_{i=1}^d z_i = 1 \right\}.$$

A barrier function for this set is

$$f_p(s, z) = \sum_{i=1}^d (-2 \log z_i - \log(z_i^{2/p} - s_i^2)).$$

Its barrier parameter is  $4 \cdot d$  (where each term in the summation has parameter 4); see Nesterov and Nemirovskii [25]. We will also use a special case of this function ( $d = 1$ ):

$$f_{1,p}(s_i, z_i) = (-2 \log z_i - \log(z_i^{2/p} - s_i^2)),$$

whose barrier parameter is 4.

Consider the conic hull

$$\begin{aligned} K(G_p) &= \left\{ (t \in R_+, s \in R^d, z \in R^d) : t > 0, \left( \frac{s}{t}, \frac{z}{t} \right) \in G_p \right\} \\ &= \left\{ (t \in R_+, s \in R^d, z \in R^d) : t > 0, t^{p-1} z_i \geq |s_i|^p, \sum_{i=1}^d z_i = t \right\}, \end{aligned}$$

which is equivalent to  $K$  for  $(t, s)$ . In what follows, we prove the following theorem.

**THEOREM 3.1.** *The function*

$$\hat{f}_p(t, s, z) = 25 \cdot \left( f_p \left( \frac{s}{t}, \frac{z}{t} \right) - 8d \log t \right)$$

*is a logarithmically homogeneous self-concordant barrier for  $K(G_p)$ , where the barrier parameter is  $200d$ .*

*Proof.* Notice that

$$\hat{f}_p(t, s, z) = 25 \cdot \left( f_p \left( \frac{s}{t}, \frac{z}{t} \right) - 8d \log t \right) = \sum_{i=1}^d 25 \cdot \left( f_{1,p} \left( \frac{s_i}{t}, \frac{z_i}{t} \right) - 8 \log t \right).$$

It is sufficient to prove that

$$\begin{aligned} \hat{f}_{1,p}(t \in R_+, s \in R, z \in R_+) &= 25 \cdot (-2 \log(z/t) - \log((z/t)^{2/p} - (s/t)^2) - 8 \log t) \\ &= 25 \cdot (-2 \log z - \log(z^{2/p} t^{2(p-1)/p} - s^2) - 4 \log t) \end{aligned}$$

is a logarithmically homogeneous self-concordant barrier function with parameter 200. Our proof follows from Proposition 5.1.4 of Nesterov and Nemirovskii [25]. Note that we have changed notation again: we have dropped the subscript  $i$  in  $s_i$  and  $z_i$  to simplify notation in the rest of this proof.

Let us fix  $o = (t, z, s)$ ,  $t > 0$ , and  $t^{p-1}z > |s|^p$ , and let  $w = (-d_t, d_z, d_s) \in R^3$ . Let us compute the derivatives up to the order 3 of  $\hat{f}_{1,p}$  at the point  $o$  in the direction  $w$ . For  $\sigma = d_t/t$  and some  $\alpha \in R$ , we have

$$\frac{s + \alpha d_s}{t - \alpha d_t} = \frac{s}{t} + \frac{\alpha}{1 - \alpha\sigma} \left( \sigma \frac{s}{t} + \frac{d_s}{t} \right)$$

and

$$\frac{z + \alpha d_z}{t - \alpha d_t} = \frac{z}{t} + \frac{\alpha}{1 - \alpha\sigma} \left( \sigma \frac{z}{t} + \frac{d_z}{t} \right).$$

Let

$$\phi(\alpha) := f_{1,p} \left( \frac{s}{t} + \alpha \left( \sigma \frac{s}{t} + \frac{d_s}{t} \right), \frac{z}{t} + \alpha \left( \sigma \frac{z}{t} + \frac{d_z}{t} \right) \right).$$

Then

$$\hat{\phi}(\alpha) := \hat{f}_{1,p}(o + \alpha w) = 25 \cdot \left( \phi \left( \frac{\alpha}{1 - \alpha\sigma} \right) - 8 \log(1 - \alpha\sigma) - 8 \log t \right).$$

Therefore, we have

$$\begin{aligned} \hat{\pi}_1 &:= \nabla \hat{f}_{1,p}(o)[w] = \hat{\phi}'(0) = 25(8\sigma + \pi_1), \\ \hat{\pi}_2 &:= \nabla^2 \hat{f}_{1,p}(o)[w, w] = \hat{\phi}''(0) = 25(8\sigma^2 + 2\sigma\pi_1 + \pi_2), \\ \hat{\pi}_3 &:= \nabla^3 \hat{f}_{1,p}(o)[w, w, w] = \hat{\phi}'''(0) = 25(16\sigma^3 + 6\sigma^2\pi_1 + 6\sigma\pi_2 + \pi_3), \end{aligned}$$

where

$$\pi_1 = \phi'(0), \quad \pi_2 = \phi''(0), \quad \pi_3 = \phi'''(0).$$

Since  $f_p$  is a self-concordant barrier with the barrier parameter 4, we have

$$\pi_2 \geq 0, \quad \pi_1^2 \leq 4\pi_2, \quad \text{and} \quad |\pi_3| \leq 2\pi_2^{3/2}.$$

Thus, we see

$$\hat{\pi}_2 \geq 25(8\sigma^2 - 4|\sigma|\sqrt{\pi_2} + \pi_2) \geq 25 \max(4\sigma^2, \pi_2/2) \geq 0,$$

which implies that  $\hat{f}_{1,p}$  is a convex function. Now we need to prove that

$$|\hat{\pi}_3| \leq 2\hat{\pi}_2^{3/2}.$$

Note that, if  $\sigma \leq 0$ , we have

$$\begin{aligned} \hat{\pi}_3/25 &= 3\sigma(8\sigma^2 + 2\sigma\pi_1 + \pi_2) - 8\sigma^3 + 3\sigma\pi_2 + \pi_3 \\ &\leq -8\sigma^3 + \pi_3 \\ &\leq 8|\sigma|^3 + 2\pi_2^{3/2} \\ &\leq \frac{8}{1000}\hat{\pi}_2^{3/2} + \frac{4\sqrt{2}}{125}\hat{\pi}_2^{3/2} \\ &\leq \frac{2}{25}\hat{\pi}_2^{3/2}, \end{aligned}$$

and

$$\begin{aligned} -\hat{\pi}_3/25 &\leq -3\sigma(8\sigma^2 + 2\sigma\pi_1 + \pi_2) - 3\sigma\pi_2 - \pi_3 \\ &\leq \frac{3}{250}\hat{\pi}_2^{3/2} - 3\sigma\pi_2 + 2\pi_2^{3/2}. \end{aligned}$$

For the latter, consider the following maximum problem for any fixed  $a \geq 0$ :

$$\max -3xy^2 + 2y^3 : 8x^2 + 4xy + y^2 = a, \quad x \leq 0, \quad y \geq 0.$$

One can verify that the maximum value is below  $8a^{3/2}$ . Thus,

$$-3\sigma\pi_2 + 2\pi_2^{3/2} \leq \frac{8}{125}\hat{\pi}_2^{3/2},$$

which, together with the above inequality, implies that

$$-\hat{\pi}_3 \leq \left(\frac{3}{10} + \frac{8}{5}\right)\hat{\pi}_2^{3/2} \leq 2\hat{\pi}_2^{3/2}.$$

Thus, for the case of  $\sigma \leq 0$ , we have

$$|\hat{\pi}_3| \leq 2\hat{\pi}_2^{3/2}.$$

Now we consider the case when  $\sigma > 0$ :

$$\begin{aligned} \hat{\pi}_3/25 &= 3\sigma(8\sigma^2 + 2\sigma\pi_1 + \pi_2) - 8\sigma^3 + 3\sigma\pi_2 + \pi_3 \\ &\leq 3\sigma(8\sigma^2 + 2\sigma\pi_1 + \pi_2) + 3\sigma\pi_2 + \pi_3 \\ &\leq \frac{3}{25}\sigma\hat{\pi}_2 + 3\sigma\pi_2 + 2\pi_2^{3/2} \\ &\leq \frac{3}{250}\hat{\pi}_2^{3/2} + 3\sigma\pi_2 + 2\pi_2^{3/2}. \end{aligned}$$

Consider again the following maximum problem for any fixed  $a > 0$ :

$$\max 3xy^2 + 2y^3 : 8x^2 - 4xy + y^2 = a, \quad x \geq 0, \quad y \geq 0,$$

where the maximum value is below  $8a^{3/2}$ . Thus,

$$3\sigma\pi_2 + 2\pi_2^{3/2} \leq \frac{8}{125}\hat{\pi}_2^{3/2},$$

which, together with the above inequality, implies that

$$\hat{\pi}_3 \leq \left(\frac{3}{10} + \frac{8}{5}\right) \hat{\pi}_2^{3/2} \leq 2\hat{\pi}_2^{3/2}.$$

Also,

$$\begin{aligned} -\hat{\pi}_3/25 &\leq 8\sigma^3 - \pi_3 \\ &\leq 8|\sigma|^3 + 2\pi_2^{3/2} \\ &\leq \frac{8}{1000} \hat{\pi}_2^{3/2} + \frac{4\sqrt{2}}{125} \hat{\pi}_2^{3/2} \\ &\leq \frac{2}{25} \hat{\pi}_2^{3/2}. \end{aligned}$$

To summarize, we have proved that

$$|\hat{\pi}_3| \leq 2\hat{\pi}_2^{3/2}.$$

Finally, it is easy to verify that the barrier parameter is 200 for  $\hat{f}_{1,p}$ . Therefore, the barrier parameter for  $\hat{f}_p$  is  $200d$ . This completes the proof of the theorem.  $\square$

**3.2. Barrier function II.** In this section, we first consider the set

$$\left\{ z \in R^d : z_i \geq 0, i = 1, \dots, d, \sum_{i=1}^d z_i^p \leq 1 \right\}.$$

A barrier function for this set is

$$f_p(z) = -4 \log \left( 1 - \sum_{i=1}^d z_i^p \right) - 4p^2 \sum_{i=1}^d \log z_i.$$

It is a convex barrier function. We present its barrier parameter in the following theorem.

**THEOREM 3.2.** *The function  $f_p(z)$  is a logarithmically self-concordant function with parameter  $4(1 + p^2d)$  in the interior of the set  $\{z \in R^d : \sum_{i=1}^d z_i^p \leq 1, z \geq 0\}$ .*

*Proof.* Let  $z > 0$  and  $\delta = \sum_{i=1}^d z_i^p < 1$  be given, and let  $w \in R^d$ . Let

$$\phi(\alpha) := f_p(z + \alpha w).$$

We compute the derivatives up to the order 3 of  $f_p$  at the point  $z$  in the direction  $w$ :

$$\begin{aligned} \pi_1/4 &:= \nabla f_p(z)[w]/4 = \phi'(0)/4 \\ &= \frac{p}{1-\delta} \sum_{i=1}^d w_i z_i^{p-1} - p^2 \sum_{i=1}^d w_i z_i^{-1}, \end{aligned}$$

$$\begin{aligned} \pi_2/4 &:= \nabla^2 f_p(z)[w, w]/4 = \phi''(0)/4 \\ &= \frac{p^2}{(1-\delta)^2} \left( \sum_{i=1}^d w_i z_i^{p-1} \right)^2 + \frac{p(p-1)}{1-\delta} \sum_{i=1}^d w_i^2 z_i^{p-2} + p^2 \sum_{i=1}^d w_i^2 z_i^{-2}, \end{aligned}$$

and

$$\pi_3/4 := \nabla^3 f_p(z)[w, w, w]/4 = \phi'''(0)/4$$



$$\begin{aligned}
 &= \frac{2p^3}{(1-\delta)^3} \left( \sum_{i=1}^d w_i z_i^{p-1} \right)^3 + \frac{2p^2(p-1)}{(1-\delta)^2} \left( \sum_{i=1}^d w_i z_i^{p-1} \right) \left( \sum_{i=1}^d w_i^2 z_i^{p-2} \right) \\
 &+ \frac{p^2(p-1)}{(1-\delta)^2} \left( \sum_{i=1}^d w_i z_i^{p-1} \right) \left( \sum_{i=1}^d w_i^2 z_i^{p-2} \right) + \frac{p(p-1)(p-2)}{1-\delta} \sum_{i=1}^d w_i^3 z_i^{p-3} + p^2 \sum_{i=1}^d w_i^3 z_i^{-3}.
 \end{aligned}$$

Now let

$$\begin{aligned}
 \phi_1 &= \frac{p}{1-\delta} \sum_{i=1}^d w_i z_i^{p-1}, & \phi_2 &= p^2 \sum_{i=1}^d w_i z_i^{-1}, \\
 \phi_3 &= \sqrt{\frac{p(p-1)}{1-\delta} \sum_{i=1}^d w_i^2 z_i^{p-2}}, & \text{and } \phi_4 &= \sqrt{p^2 \sum_{i=1}^d w_i^2 z_i^{-2}}.
 \end{aligned}$$

Then,

$$\begin{aligned}
 \pi_1/4 &= \phi_1 - \phi_2, \\
 \pi_2/4 &= \phi_1^2 + \phi_3^2 + \phi_4^2 > 0,
 \end{aligned}$$

and

$$\begin{aligned}
 |\pi_3/4| &\leq |2\phi_1^3 + 3\phi_1\phi_3^2| + \frac{|p-2|}{p} \phi_3^2\phi_4 + \frac{1}{p} \phi_4^3 \\
 &\leq 2|\phi_1|(\phi_1^2 + \phi_3^2) + |\phi_1|\phi_3^2 + \phi_3(\phi_3^2 + \phi_4^2) \\
 &\leq \pi_2^{3/2}/4 + \pi_2^{3/2}/8 + \pi_2^{3/2}/8 \\
 &= \pi_2^{3/2}/2;
 \end{aligned}$$

i.e.,

$$|\pi_3| \leq 2\pi_2^{3/2}.$$

Finally, we need to prove

$$\pi_1^2 \leq 4(1 + p^2d)\pi_2.$$

Note that

$$\pi_1^2 = 16(\phi_1 - \phi_2)^2.$$

Thus, if  $\phi_1 = 0$ , then

$$\pi_1^2 = 16\phi_2^2 \leq 16p^2d\phi_4^2 \leq 4p^2d\pi_2.$$

Consider  $\phi_1 \neq 0$  and let  $\bar{\delta} = \frac{|\phi_2|}{|\phi_1|}$ . Then

$$\pi_2/4 \geq \phi_1^2 + \frac{\phi_2^2}{p^2d} = \left(1 + \frac{\bar{\delta}^2}{p^2d}\right) \phi_1^2.$$

On the other hand,

$$\pi_1^2 \leq 16(1 + \bar{\delta})^2 \phi_1^2.$$

Therefore,

$$\begin{aligned} \pi_1^2 &\leq \frac{(1 + \bar{\delta})^2}{1 + \frac{\bar{\delta}^2}{p^2 d}} \cdot 4\pi_2 \\ &\leq 4(1 + p^2 d)\pi_2, \end{aligned}$$

where the last inequality holds since

$$(1 + \bar{\delta})^2 \leq (1 + p^2 d) \left(1 + \frac{\bar{\delta}^2}{p^2 d}\right). \quad \square$$

Similarly, we can prove the following corollary.

**COROLLARY 3.3.** *If  $p \leq 3$ , then*

$$f_p(z) = -4 \log \left(1 - \sum_{i=1}^d z_i^p\right) - 4 \sum_{i=1}^d \log z_i$$

*is a convex, logarithmically self-concordant barrier function with parameter  $4(1 + d)$  in the interior of the set  $\{z \in R^d : \sum_{i=1}^d z_i^p \leq 1, z \geq 0\}$ .*  $\square$

We now consider the set

$$G_p = \left\{ (s \in R^d, z \in R^d) : z_i \geq |s_i|, i = 1, \dots, d, \sum_{i=1}^d z_i^p \leq 1 \right\}.$$

A barrier function for this set is

$$f_p(s, z) = -4 \log \left(1 - \sum_{i=1}^d z_i^p\right) - 4p^2 \sum_{i=1}^d \log z_i - \sum_{i=1}^d \log(z_i^2 - s_i^2).$$

The barrier parameter of this function is  $4(1 + p^2 d) + 2d$ , since the first two summations have parameter  $4(1 + p^2 d)$  and the last summation has parameter  $2d$ .

Again consider the conic hull

$$\begin{aligned} K(G_p) &= \left\{ (t, s \in R^d, z \in R^d) : t > 0, \left(\frac{s}{t}, \frac{z}{t}\right) \in G_p \right\} \\ &= \left\{ (t, s \in R^d, z \in R^d) : t > 0, z_i \geq |s_i|, \sum_{i=1}^d z_i^p \leq t^p \right\}, \end{aligned}$$

which is also equivalent to  $K$  for  $(t, s)$ . Again, similar to Theorem 3.1, we can prove the following theorem.

**THEOREM 3.4.** *The function*

$$\hat{f}_p(t, s, z) = \theta^2 \cdot \left( f_p \left( \frac{s}{t}, \frac{z}{t} \right) - (8(1 + p^2 d) + 4d) \log t \right)$$

*is a logarithmically homogeneous self-concordant barrier for  $K(G_p)$ , where the barrier parameter is  $\theta^2(8(1 + p^2 d) + 4d)$ , where  $\theta$  is a positive constant, say,  $\theta = 5$ .*

We can simplify the above barrier function further. Note that  $z_i \geq |s_i|$  implies  $z_i \geq 0$ , and the Hessian of  $-\log(z_i^2 - s_i^2)$  is

$$H(z_i, s_i) = \frac{2}{z_i^2 - s_i^2} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{4}{(z_i^2 - s_i^2)^2} \begin{pmatrix} z_i^2 & -z_i s_i \\ -z_i s_i & s_i^2 \end{pmatrix}.$$

For  $z_i^2 > s_i^2$ , we have

$$H(z_i, s_i) - \begin{pmatrix} z_i^{-2} & 0 \\ 0 & 0 \end{pmatrix} = \frac{2}{(z_i^2 - s_i^2)^2} \begin{pmatrix} z_i^2 + s_i^2 - \frac{(z_i^2 - s_i^2)^2}{2z_i^2} & -2z_i s_i \\ -2z_i s_i & z_i^2 + s_i^2 \end{pmatrix}.$$

We can show, from  $(z_i^2 - s_i^2)^3 > 0$ , that the two-by-two matrix in the right-hand-side expression is positive semidefinite. Thus, we have the following corollary, whose proof is similar to that for Theorem 3.2.

**COROLLARY 3.5.** *If  $p \leq 3$ , then*

$$f_p(s, z) = -4 \log \left( 1 - \sum_{i=1}^d z_i^p \right) - 4 \sum_{i=1}^d \log(z_i^2 - s_i^2)$$

*is a convex, logarithmically self-concordant barrier function with parameter  $4(1 + 2d)$  in the interior of the set  $\{(s \in R^d, z \in R^d) : \sum_{i=1}^d z_i^p \leq 1, z \geq |s_i|\}$ . The function*

$$\begin{aligned} \hat{f}_p(t, s, z) &= 25 \cdot \left( f_p \left( \frac{s}{t}, \frac{z}{t} \right) - 8(1 + 2d) \log t \right) \\ &= 100 \left( -\log \left( t^p - \sum_{i=1}^d z_i^p \right) - \sum_{i=1}^d \log(z_i^2 - s_i^2) - (2 - p + 2d) \log t \right) \end{aligned}$$

*is a logarithmically homogeneous self-concordant barrier for  $K(G_p)$ , where the barrier parameter is  $200(1 + 2d)$ .  $\square$*

Symmetrically, we may consider the convex set

$$G_q = \left\{ (x \in R^d, v \in R^d) : v_i \geq |x_i|^q, i = 1, \dots, d, \sum_{i=1}^d v_i = 1 \right\}$$

and the conic hull

$$\begin{aligned} K(G_q) &= \left\{ (\tau, x \in R^d, v \in R^d) : \tau > 0, \left( \frac{x}{\tau}, \frac{v}{\tau} \right) \in G_q \right\} \\ &= \left\{ (\tau, x \in R^d, v \in R^d) : \tau > 0, v_i \geq |x_i|, \sum_{i=1}^d v_i^q \leq \tau^q \right\}, \end{aligned}$$

which is also equivalent to  $K^*$  for  $(\tau, x)$ . A direct application of Corollary 3.5 leads to the following corollary.

**COROLLARY 3.6.** *If  $q \leq 3$ , then*

$$f_q(x, v) = -4 \log \left( 1 - \sum_{i=1}^d v_i^q \right) - 4 \sum_{i=1}^d \log(v_i^2 - x_i^2)$$

*is a convex, logarithmically self-concordant barrier function with parameter  $4(1 + 2d)$  in the interior of the set  $\{(x \in R^d, v \in R^d) : \sum_{i=1}^d v_i^q \leq 1, v \geq |x_i|\}$ . The function*

$$\hat{f}_q(\tau, x, v) = 25 \cdot \left( f_q \left( \frac{x}{\tau}, \frac{v}{\tau} \right) - 8(1 + 2d) \log \tau \right)$$

$$= 100 \left( -\log \left( \tau^q - \sum_{i=1}^d v_i^q \right) - \sum_{i=1}^d \log(v_i^2 - x_i^2) - (2 - q + 2d) \log \tau \right)$$

is a logarithmically homogeneous self-concordant barrier for  $K(G_q)$ , where the barrier parameter is  $200(1 + 2d)$ .  $\square$

**3.3. Legendre transformations.** Although it has a higher barrier parameter, barrier function  $\Pi$  possesses a structure similar to the barrier function for the second-order cone, which we will use in the following analyses. Since we can solve both (P) and (D) using either the barrier function for the  $p$ -order cone in (D) or the barrier function for the  $q$ -order cone in (P), we will make the following conventions in this paper:

- When  $p > 3$ , use the barrier function of the  $q$ -order cone in (P).
- When  $1 \leq p \leq 3$ , use the barrier function of the  $p$ -order cone in (D).

It follows from Nesterov and Nemirovskii [25] that the Legendre transformation of the logarithmically homogeneous self-concordant barrier function in Corollary 3.5,

$$\hat{f}_p^*(\tau, x \in R^d) = \sup\{-\tau \cdot t - x^T s - \hat{f}_p(t, s, z) : (t, s, z) \in K(G_p)\},$$

is a logarithmically homogeneous self-concordant barrier for  $K^*$  with the same parameter  $200(1 + 2d)$ . It seems hard to find an explicit form of  $\hat{f}_p^*(\tau, x)$ . Fortunately, for the interior point algorithm presented in the next section, we do not need such an explicit formula.

We need only the initial barrier function value at a primal initial point in our complexity analysis. We will set the initial point  $\tau = 1$  and  $x = 0$ . Thus, we need to evaluate

$$\hat{f}_p^*(1, 0) = \sup\{-t - \hat{f}_p(t, s, z) : (t, s, z) \in K(G_p)\}.$$

This is a convex optimization problem with an analytical solution

$$t^* = 200(1 + 2d), \quad s^* = 0, \quad z_j^* = \left( \frac{2}{p + 2d} \right)^{1/p} t^*, \quad j = 1, \dots, d.$$

Thus,

(3.1)

$$\hat{f}_p^*(1, 0) = 200(1 + 2d)(\log(200(1 + 2d)) - 1) + 100 \log \frac{p}{p + 2d} + \frac{200d}{p} \log \frac{2}{p + 2d}.$$

It also follows from Nesterov and Nemirovskii [25] that the Legendre transformation of the logarithmically homogeneous self-concordant barrier function in Corollary 3.6,

$$\hat{f}_q^*(t, s \in R^d) = \sup\{-\tau \cdot t - x^T s - \hat{f}_q(\tau, x, v) : (\tau, x, v) \in K(G_q)\},$$

is a logarithmically homogeneous self-concordant barrier for  $K$  with the same parameter  $200(1 + 2d)$ . It seems hard to find the exact value of  $\hat{f}_q^*(t, s)$  for  $s \neq 0$ . In the following, we will find an explicit formula for  $\hat{f}_q^*(c, 0)$  for any  $c > 0$  and prove that  $\hat{f}_q^*(t, s) \leq \hat{f}_q^*(t - \|s\|_p, 0)$  for  $(t, s) \in K$ .

Note that

$$\hat{f}_q^*(c, 0) = \sup\{-\tau c - \hat{f}_q(\tau, x, v) : (\tau, x, v) \in K(G_q)\}.$$

This is a convex optimization problem with an analytical solution

$$\tau^* = 200(1 + 2d)/c, \quad x^* = 0, \quad v_j^* = \left(\frac{2}{q + 2d}\right)^{1/q} \tau^*, \quad j = 1, \dots, d.$$

Thus,

(3.2)

$$\hat{f}_q^*(c, 0) = 200(1 + 2d)(\log(200(1 + 2d)/c) - 1) + 100 \log \frac{q}{q + 2d} + \frac{200d}{q} \log \frac{2}{q + 2d}.$$

For any  $(t, s) \in K$  and  $(\tau, x) \in K^*$ , we have

$$t \geq \|s\|_p, \quad \tau \geq \|x\|_q.$$

Since

$$|s^T x| \leq \|s\|_p \|x\|_q,$$

we have

$$\begin{aligned} -t\tau - s^T x &\leq -t\tau + \|s\|_p \|x\|_q \\ &\leq -t\tau + \|s\|_p \tau \\ &= -\tau(t - \|s\|_p). \end{aligned}$$

Therefore,

(3.3) 
$$\hat{f}_q^*(t, s) \leq \hat{f}_q^*(t - \|s\|_p, 0).$$

**4. A primal-dual potential reduction algorithm.** In this section, we will present a primal-dual potential reduction algorithm for computing  $\epsilon$ -optimal solutions for  $(P)$  and  $(D)$  in polynomial time. We will use either dual scaling or primal scaling depending on the value of  $p$ .

**4.1. Use dual scaling when  $p \in [1, 3]$ .** When  $p \in [1, 3]$ , we may solve  $(P)$  and  $(D)$  using the barrier function for the  $p$ -order cone. Let

(4.1) 
$$F_p^*(\mathcal{X}) = \sum_{i=1}^m \hat{f}_p^*(\tau_i, x_i) \quad \text{and} \quad F_p(\mathcal{S}, \mathcal{Z}) = \sum_{i=1}^m \hat{f}_p(t_i, s_i, z_i \in R^d),$$

where

$$\mathcal{Z} := (z_1; z_2; \dots; z_m).$$

They are logarithmically homogeneous self-concordant barriers for  $(P)$  and  $(D)$ , where the barrier parameter is  $\theta := 200(1 + 2d)m$ . A primal-dual potential function for the pair  $(P)$  and  $(D)$  is

(4.2) 
$$\phi_\rho(\mathcal{X}, \mathcal{S}, \mathcal{Z}) := \rho \log \langle \mathcal{X}, \mathcal{S} \rangle + F_p^*(\mathcal{X}) + F_p(\mathcal{S}, \mathcal{Z}) + \theta,$$

where  $\rho = \theta + \gamma\sqrt{\theta}$ ,  $\gamma \geq 1$ . Note that

$$\langle \mathcal{X}, \mathcal{S} \rangle = \mathcal{X}^T \mathcal{S} = \mathcal{C}^T \mathcal{X} - \mathcal{B}^T \mathcal{Y},$$

and from Nesterov and Nemirovskii [25],

(4.3) 
$$\phi_\theta(\mathcal{X}, \mathcal{S}, \mathcal{Z}) \geq \theta \log \theta.$$

The main iteration of a potential reduction algorithm starts with a strictly feasible primal-dual pair  $\mathcal{X}$  and  $(\mathcal{Y}, \mathcal{S})$ , i.e.,

$$\begin{aligned} \mathcal{A}\mathcal{X} &= \mathcal{B}, & \mathcal{S} &= \mathcal{C} - \mathcal{A}^T\mathcal{Y}, \\ \mathcal{X} &\in \text{int}\mathcal{K}^*, & \text{and } \mathcal{S} &\in \text{int}\mathcal{K}. \end{aligned}$$

It computes a search direction  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{Z}})$  by solving a system of linear equations. After obtaining  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{Z}})$ , a new, strictly feasible, primal-dual pair  $\mathcal{X}^+$  and  $(\mathcal{Y}^+, \mathcal{S}^+)$  is generated from

$$\mathcal{X}^+ = \mathcal{X} + \alpha d_{\mathcal{X}}, \quad \mathcal{Y}^+ = \mathcal{Y} + \beta d_{\mathcal{Y}}, \quad \mathcal{S}^+ = \mathcal{S} + \beta d_{\mathcal{S}}, \quad \mathcal{Z}^+ = \mathcal{Z} + \beta d_{\mathcal{Z}},$$

for some step-sizes  $\alpha$  and  $\beta$ , and

$$\phi_{\rho}(\mathcal{X}^+, \mathcal{S}^+, \mathcal{Z}^+) \leq \phi_{\rho}(\mathcal{X}, \mathcal{S}, \mathcal{Z}) - \Omega(1).$$

The search direction  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{Z}})$  is determined by the following equations:

$$(4.4) \quad \mathcal{A}d_{\mathcal{X}} = 0, \quad d_{\mathcal{S}} = -\mathcal{A}^T d_{\mathcal{Y}} \quad (\text{feasibility})$$

and

$$(4.5) \quad \begin{pmatrix} d_{\mathcal{X}} \\ 0 \end{pmatrix} + F_p''(\mathcal{S}, \mathcal{Z}) \begin{pmatrix} d_{\mathcal{S}} \\ d_{\mathcal{Z}} \end{pmatrix} = -\frac{\rho}{\mathcal{X}^T \mathcal{S}} \begin{pmatrix} \mathcal{X} \\ 0 \end{pmatrix} - F_p'(\mathcal{S}, \mathcal{Z}).$$

The *theoretical* dual-scaling potential reduction algorithm, with a specific choice of step-sizes  $\alpha$  and  $\beta$ , can be described as follows (see section 4.5.3 of [25]).

ALGORITHM PDD

Let  $\gamma$  and  $\Delta$  be fixed constants such that  $\gamma \geq 1$ ,  $0 < \Delta < 1$ , and  $\frac{\gamma(\gamma(1-\Delta)-\Delta)}{1+\gamma} > \frac{\Delta^2}{2(1-\Delta)^2}$ .

Step 1. Compute the search direction  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{Z}})$  using (4.4) and (4.5).

Step 2. Compute  $\lambda = \sqrt{\begin{pmatrix} d_{\mathcal{S}} \\ d_{\mathcal{Z}} \end{pmatrix}^T F_p''(\mathcal{S}, \mathcal{Z}) \begin{pmatrix} d_{\mathcal{S}} \\ d_{\mathcal{Z}} \end{pmatrix}}$ .

If  $\lambda > \Delta$ , then

$$\mathcal{X}^+ = \mathcal{X} \quad (\text{primal step-size } \alpha = 0),$$

$$\mathcal{S}^+ = \mathcal{S} + \frac{1}{1+\lambda} d_{\mathcal{S}} \quad (\text{dual step-size } \beta = \frac{1}{1+\lambda}),$$

$$\mathcal{Y}^+ = \mathcal{Y} + \frac{1}{1+\lambda} d_{\mathcal{Y}},$$

$$\mathcal{Z}^+ = \mathcal{Z} + \frac{1}{1+\lambda} d_{\mathcal{Z}},$$

else

$$\mathcal{X}^+ = \mathcal{X} + \frac{\langle \mathcal{S}, \mathcal{X} \rangle}{\rho} d_{\mathcal{X}} \quad (\text{primal step-size } \alpha = \frac{\langle \mathcal{S}, \mathcal{X} \rangle}{\rho}),$$

$$\mathcal{S}^+ = \mathcal{S} \quad (\text{dual step-size } \beta = 0),$$

$$\mathcal{Y}^+ = \mathcal{Y},$$

$$\mathcal{Z}^+ = \mathcal{Z}.$$

endif

According to Nesterov and Nemirovskii [25], we have the following theorem.

**THEOREM 4.1.** *Starting from any strictly feasible primal solution  $\mathcal{X}^0$  and strictly dual feasible solution  $(\mathcal{Y}^0; \mathcal{S}^0)$ , an  $\epsilon$ -optimal solution  $(\mathcal{X}, \mathcal{Y}, \mathcal{S})$  to problem (1.1), can be obtained by repeated explication of Algorithm PDD for at most  $O(\gamma\sqrt{\theta} \log(\langle \mathcal{X}^0, \mathcal{S}^0 \rangle / \epsilon)) + \phi_{\theta}(\mathcal{X}^0, \mathcal{S}^0, \mathcal{Z}^0) - \theta \log \theta$  iterations.  $\square$*

In practice, one usually finds the largest step-sizes  $\bar{\alpha}$  and  $\bar{\beta}$  such that

$$(4.6) \quad \mathcal{X} + \bar{\alpha}d_{\mathcal{X}} \in \mathcal{K}^* \quad \text{and} \quad \mathcal{S} + \bar{\beta}d_{\mathcal{S}} \in \mathcal{K},$$

then takes  $\alpha \in [0, \bar{\alpha}]$  and  $\beta \in [0, \bar{\beta}]$  via a line-search to minimize  $\phi_{\rho}(\mathcal{X}^+, \mathcal{S}^+, \mathcal{Z}^+)$ , or simply chooses

$$(4.7) \quad \alpha = (0.5 \sim 0.99)\bar{\alpha} \quad \text{and} \quad \beta = (0.5 \sim 0.99)\bar{\beta}$$

as long as  $\phi_{\rho}$  is reduced.

**4.2. Use primal scaling when  $p \in [3, \infty]$ .** The dual scaling algorithm is good when  $p$  is small ( $1 \leq p \leq 3$ ). For larger values of  $p$ , the barrier parameter for the  $p$ -order cone becomes larger. In this case, it is better to use the barrier function of the  $q$ -order cone in  $(P)$ , where  $\frac{1}{p} + \frac{1}{q} = 1$ . It is clear that  $q \in [1, 2]$  whenever  $p \in [2, \infty]$ .

Let

$$(4.8) \quad F_q(\mathcal{X}, \mathcal{V}) = \sum_{i=1}^m \hat{f}_q(\tau_i, x_i, v_i \in R^d) \quad \text{and} \quad F_q^*(\mathcal{S}) = \sum_{i=1}^m \hat{f}_q^*(t_i, s_i),$$

where

$$\mathcal{V} := (v_1; v_2; \dots; v_m).$$

They are logarithmically homogeneous self-concordant barriers for  $(P)$  and  $(D)$ , where the barrier parameter is  $\theta := 200(1 + 2d)m$ . A primal-dual potential function for the pair  $(P)$  and  $(D)$  is

$$(4.9) \quad \psi_{\rho}(\mathcal{X}, \mathcal{S}, \mathcal{V}) := \rho \log \langle \mathcal{X}, \mathcal{S} \rangle + F_q(\mathcal{X}, \mathcal{V}) + F_q^*(\mathcal{S}) + \theta,$$

where  $\rho = \theta + \gamma\sqrt{\theta}$ ,  $\gamma \geq 1$ . Note that

$$\langle \mathcal{X}, \mathcal{S} \rangle = \mathcal{X}^T \mathcal{S} = \mathcal{C}^T \mathcal{X} - \mathcal{B}^T \mathcal{Y},$$

and from Nesterov and Nemirovskii [25]

$$(4.10) \quad \psi_{\theta}(\mathcal{X}, \mathcal{S}, \mathcal{V}) \geq \theta \log \theta.$$

The main iteration of a potential reduction algorithm starts with a strictly feasible primal-dual pair  $(\mathcal{X}, \mathcal{V})$  and  $(\mathcal{Y}, \mathcal{S})$ , i.e.,

$$\begin{aligned} \mathcal{A}\mathcal{X} &= \mathcal{B}, & \mathcal{S} &= \mathcal{C} - \mathcal{A}^T \mathcal{Y}, \\ \mathcal{X} &\in \text{int}\mathcal{K}, & \mathcal{S} &\in \text{int}\mathcal{K}^*. \end{aligned}$$

It computes a search direction  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{V}})$  by solving a system of linear equations. After obtaining  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{V}})$ , a new, strictly feasible, primal-dual pair  $\mathcal{X}^+$  and  $(\mathcal{Y}^+, \mathcal{S}^+)$  is generated from

$$\mathcal{X}^+ = \mathcal{X} + \alpha d_{\mathcal{X}}, \quad \mathcal{Y}^+ = \mathcal{Y} + \beta d_{\mathcal{Y}}, \quad \mathcal{S}^+ = \mathcal{S} + \beta d_{\mathcal{S}}, \quad \mathcal{V}^+ = \mathcal{V} + \alpha d_{\mathcal{V}}$$

for some step-sizes  $\alpha$  and  $\beta$ , and

$$\psi_{\rho}(\mathcal{X}^+, \mathcal{S}^+, \mathcal{V}^+) \leq \psi_{\rho}(\mathcal{X}, \mathcal{S}, \mathcal{V}) - \Omega(1).$$

The search direction  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{V}})$  is determined by (4.4) and the following equations:

$$(4.11) \quad \begin{pmatrix} d_{\mathcal{S}} \\ 0 \end{pmatrix} + F_q''(\mathcal{X}, \mathcal{V}) \begin{pmatrix} d_{\mathcal{X}} \\ d_{\mathcal{V}} \end{pmatrix} = -\frac{\rho}{\mathcal{X}^T \mathcal{S}} \begin{pmatrix} \mathcal{S} \\ 0 \end{pmatrix} - F_q'(\mathcal{X}, \mathcal{V}).$$

The algorithm generates an  $\epsilon$ -optimal solution  $(\mathcal{X}, \mathcal{S}, \mathcal{V})$ , i.e.,

$$\langle \mathcal{X}, \mathcal{S} \rangle \leq \epsilon,$$

in a guaranteed  $O(\gamma\sqrt{\theta} \log(\langle \mathcal{X}^0, \mathcal{S}^0 \rangle / \epsilon) + \psi_\theta(\mathcal{X}^0, \mathcal{S}^0, \mathcal{V}^0) - \theta \log \theta)$  iterations.

In practice, one usually finds the largest step-sizes  $\bar{\alpha}$  and  $\bar{\beta}$  such that

$$(4.12) \quad \mathcal{X} + \bar{\alpha}d_{\mathcal{X}} \in \mathcal{K} \quad \text{and} \quad \mathcal{S} + \bar{\beta}d_{\mathcal{S}} \in \mathcal{K}^*,$$

then takes  $\alpha \in [0, \bar{\alpha}]$  and  $\beta \in [0, \bar{\beta}]$  via a line-search to minimize  $\psi_\rho(\mathcal{X}^+, \mathcal{S}^+, \mathcal{V}^+)$ , or simply chooses

$$(4.13) \quad \alpha = (0.5 \sim 0.99)\bar{\alpha} \quad \text{and} \quad \beta = (0.5 \sim 0.99)\bar{\beta}$$

as long as  $\psi_\rho$  is reduced.

The *theoretical* potential reduction algorithm using primal scaling can be described as follows.

ALGORITHM PDP.

Let  $\gamma$  and  $\Delta$  be fixed constants such that  $\gamma \geq 1$ ,  $0 < \Delta < 1$ , and  $\frac{\gamma(\gamma(1-\Delta)-\Delta)}{1+\gamma} > \frac{\Delta^2}{2(1-\Delta)^2}$ .

Step 1. Compute the search direction  $(d_{\mathcal{X}}, d_{\mathcal{Y}}, d_{\mathcal{S}}, d_{\mathcal{V}})$  using (4.4) and (4.11).

Step 2. Compute  $\lambda = \sqrt{\begin{pmatrix} d_{\mathcal{X}} \\ d_{\mathcal{V}} \end{pmatrix}^T F''_q(\mathcal{X}, \mathcal{V}) \begin{pmatrix} d_{\mathcal{X}} \\ d_{\mathcal{V}} \end{pmatrix}}$ .

If  $\lambda > \Delta$ , then

$$\mathcal{X}^+ = \mathcal{X} + \frac{1}{1+\lambda}d_{\mathcal{X}} \quad (\text{primal step-size } \alpha = \frac{1}{1+\lambda}),$$

$$\mathcal{V}^+ = \mathcal{V} + \frac{1}{1+\lambda}d_{\mathcal{V}},$$

$$\mathcal{S}^+ = \mathcal{S} \quad (\text{dual step-size } \beta = 0),$$

$$\mathcal{Y}^+ = \mathcal{Y},$$

else

$$\mathcal{X}^+ = \mathcal{X} \quad (\text{primal step-size } \alpha = 0),$$

$$\mathcal{V}^+ = \mathcal{V},$$

$$\mathcal{S}^+ = \mathcal{S} + \frac{\langle \mathcal{S}, \mathcal{X} \rangle}{\rho}d_{\mathcal{S}} \quad (\text{dual step-size } \beta = \frac{\langle \mathcal{S}, \mathcal{X} \rangle}{\rho}),$$

$$\mathcal{Y}^+ = \mathcal{Y} + \frac{\langle \mathcal{S}, \mathcal{X} \rangle}{\rho}d_{\mathcal{Y}}.$$

endif

According to Nesterov and Nemirovskii [25], we have the following theorem.

**THEOREM 4.2.** *Starting from any strictly feasible primal solution  $(\mathcal{X}^0; \mathcal{V}^0)$  and strictly dual feasible solution  $(\mathcal{Y}^0; \mathcal{S}^0)$ , an  $\epsilon$ -optimal solution to problem (1.1) can be obtained by repeated application of Algorithm PDP for at most  $O(\gamma\sqrt{\theta} \log(\langle \mathcal{X}^0, \mathcal{S}^0 \rangle / \epsilon) + \psi_\theta(\mathcal{X}^0, \mathcal{S}^0, \mathcal{V}^2) - \theta \log \theta)$  iterations.  $\square$*

**5. Complexity and implementation.** As we have seen, the number of iterations required to compute an  $\epsilon$ -optimal solution to problem (2.1) depends on the initial point  $(\mathcal{X}^0, \mathcal{S}^0, \mathcal{Z}^0)$ . In this section, we discuss the initial point selection and other computational issues for solving problem (2.1) using the algorithms presented in section 4.

**5.1. Initial point for dual scaling.** The algorithms discussed in the previous section all require a pair of strictly primal-dual interior feasible solutions. In the following, we give one such pair.



Let

$$\bar{c} = \max_{1 \leq i \leq m} \|c_i\|,$$

and for  $i = 1, 2, \dots, m$ , let

$$\begin{aligned} u^0 &= 0, & t_i^0 &= \sqrt{\|c_i\|^2 + m(1+2d)\bar{c}^2} \left(\frac{p+2d}{2}\right)^{\frac{1}{p}}, \\ s_i^0 &= c_i, & z_i^0 &= \sqrt{\|c_i\|^2 + m(1+2d)\bar{c}^2} e \in R^d, \end{aligned}$$

and

$$\tau_i^0 = 1, \quad x_i^0 = 0 \in R^d.$$

Then, one can verify that  $\mathcal{X}^0$  is an interior feasible solution to (P) and that  $\mathcal{S}^0$  and  $\mathcal{Y}^0$  form an interior feasible solution to (D). One can also verify that

$$\begin{aligned} \langle \mathcal{X}^0, \mathcal{S}^0 \rangle &= (\mathcal{X}^0)^T \mathcal{S}^0 = \sum_{i=1}^m t_i^0 \tau_i^0 = \left(\frac{p+2d}{2}\right)^{\frac{1}{p}} \sum_{i=1}^m \sqrt{\|c_i\|^2 + m(1+2d)\bar{c}^2} \\ &\leq \left(\frac{p+2d}{2}\right)^{\frac{1}{p}} \bar{c} m \sqrt{1+m(1+2d)}, \end{aligned}$$

and the initial value

$$\begin{aligned} \hat{f}_p(t_i^0, s_i^0, z_i^0) &\leq 100 \left( -(2+2d) \log t_i^0 - \log \frac{p}{p+2d} - d \log m(1+2d)\bar{c}^2 \right) \\ &\leq -100(1+2d) \log m(1+2d)\bar{c}^2 - \frac{200(1+d)}{p} \log \frac{p+2d}{2} - 100 \log \frac{p}{p+2d}. \end{aligned}$$

From this inequality and (3.1), we have

$$\begin{aligned} &F_p^*(\mathcal{X}^0) + F_p(\mathcal{S}^0, \mathcal{Z}^0) \\ &= \sum_{i=1}^m \hat{f}_p^*(1, 0) + \sum_{i=1}^m \hat{f}_p(t_i^0, s_i^0, z_i^0) \\ &\leq \theta(\log(200(1+2d)) - 1) - 100(1+2d)m \log m(1+2d)\bar{c}^2 + \frac{\theta}{p} \log \frac{2}{p+2d}. \end{aligned}$$

Thus, from these inequalities,

$$\begin{aligned} &\phi_\theta(\mathcal{X}^0, \mathcal{S}^0, \mathcal{Y}^0) - \theta \log \theta \\ &= \theta \log \langle \mathcal{X}^0, \mathcal{S}^0 \rangle + F_p^*(\mathcal{X}^0) + F_p(\mathcal{S}^0, \mathcal{Z}^0) + \theta - \theta \log \theta \\ &\leq \theta \log(\bar{c} m \sqrt{1+m(1+2d)}) - 100(1+2d)m \log m(1+2d)\bar{c}^2 \\ &\quad + \theta \log(200(1+2d)) - \theta \log \theta \\ &= 100(1+2d)m \log(1+m(1+2d)) - 100(1+2d)m \log m(1+2d) \\ &= 100(1+2d)m \log \left( 1 + \frac{1}{m(1+2d)} \right) \\ &\leq 100. \end{aligned}$$

With this initial point, we have the following corollary.

COROLLARY 5.1. *Let the initial feasible primal solution  $\mathcal{X}^0$  and dual feasible solution  $(\mathcal{Y}^0, \mathcal{S}^0, \mathcal{Z}^0)$  be given as above, and let  $1 \leq p \leq 3$ . Then, an  $\epsilon$ -optimal solution to problem (2.1) can be obtained by the (dual) potential reduction algorithm in at most*

$$O\left(\gamma\sqrt{200(1+2d)m}(\log(\bar{c}/\epsilon) + \log(md))\right)$$

iterations, where

$$\bar{c} = \max_{1 \leq i \leq m} \|c_i\|. \quad \square$$

**5.2. Initial point for primal scaling.** In this section, we give a pair of strictly primal-dual interior feasible solutions for the primal scaling algorithm. It is assumed that  $p \geq 2$ .

Let

$$\bar{c} = \max_{1 \leq i \leq m} \|c_i\|,$$

and for  $i = 1, 2, \dots, m$ , let

$$u^0 = 0, \quad s_i^0 = c_i, \quad t_i^0 = \|c_i\|_p + m(1 + 2d)\bar{c}$$

and

$$\tau_i^0 = 1, \quad x_i^0 = 0 \in R^d, \quad v_i^0 = \left(\frac{2}{q + 2d}\right)^{\frac{1}{q}} e.$$

Then, one can verify that  $\mathcal{X}^0$  is an interior feasible solution to (P) and that  $\mathcal{S}^0$  and  $\mathcal{Y}^0$  form an interior feasible solution to (D). One can also verify that

$$\langle \mathcal{X}^0, \mathcal{S}^0 \rangle = (\mathcal{X}^0)^T \mathcal{S}^0 = \sum_{i=1}^m t_i^0 \tau_i^0 \leq m\bar{c} + m^2(1 + 2d)\bar{c},$$

and the initial value

$$\begin{aligned} \hat{f}_q(\tau_i^0, x_i^0, v_i^0) &= 100 \left( -\log \left( (\tau_i^0)^q - \sum_{j=1}^d (v_j^0)^q \right) - \sum_{j=1}^d \log(v_j^0)^2 - (2 - q + 2d) \log \tau_i^0 \right) \\ &= 100 \left( -\log \left( 1 - \frac{2d}{q + 2d} \right) - \frac{2d}{q} \log \frac{2}{q + 2d} \right) \\ &= 100 \left( -\log \left( \frac{q}{q + 2d} \right) - \frac{2d}{q} \log \frac{2}{q + 2d} \right), \\ \hat{f}_i^*(t_i, s_i) &\leq 200(1 + 2d) \left( \log \frac{200(1 + 2d)}{m(1 + 2d)\bar{c}} - 1 \right) \\ &\quad + 100 \left( \log \frac{q}{q + 2d} + \frac{2d}{q} \log \frac{2}{q + 2d} \right). \end{aligned}$$

Therefore,

$$\hat{f}_q(\tau_i, x_i, v_i) + \hat{f}_i^*(t_i, s_i) \leq 200(1 + 2d) \left( \log \frac{200(1 + 2d)}{m(1 + 2d)\bar{c}} - 1 \right).$$

From these inequalities, we have

$$\begin{aligned} & F_q^*(\mathcal{S}^0) + F_q(\mathcal{X}^0, \mathcal{V}^0) \\ &= \sum_{i=1}^m \hat{f}_q^*(t_i^0, s_i^0) + \sum_{i=1}^m \hat{f}_q(\tau_i^0, x_i^0, v_i^0) \\ &\leq \theta \left( \log \frac{200(1+2d)}{m(1+2d)\bar{c}} - 1 \right). \end{aligned}$$

Thus,

$$\begin{aligned} & \psi_\theta(\mathcal{X}^0, \mathcal{S}^0, \mathcal{V}^0) - \theta \log \theta \\ &= \theta \log \langle \mathcal{X}^0, \mathcal{S}^0 \rangle + F_q^*(\mathcal{S}^0) + F_q(\mathcal{X}^0, \mathcal{V}^0) + \theta - \theta \log \theta \\ &\leq \theta \log(m\bar{c} + m^2(1+2d)\bar{c}) + \theta \left( \log \frac{200(1+2d)}{m(1+2d)\bar{c}} - 1 \right) + \theta - \theta \log \theta \\ &= \theta \log \left( 1 + \frac{1}{m(1+2d)} \right) \\ &\leq 200. \end{aligned}$$

With this initial point, we have the following corollary.

**COROLLARY 5.2.** *Let the initial feasible primal solution  $(\mathcal{X}^0, \mathcal{V}^0)$  and dual feasible solution  $(\mathcal{Y}^0, \mathcal{S}^0)$  be given as above, and let  $2 \leq q \leq 3$ . Then, an  $\epsilon$ -optimal solution to problem (2.1) can be obtained by the (primal) potential reduction algorithm in at most*

$$O \left( \gamma \sqrt{200(1+2d)m} (\log(\bar{c}/\epsilon) + \log(md)) \right)$$

iterations, where

$$\bar{c} = \max_{1 \leq i \leq m} \|c_i\|. \quad \square$$

**5.3. Search direction.** At each step of the potential reduction algorithm, we need to compute the search direction  $d_{\mathcal{X}}$ ,  $d_{\mathcal{S}}$ ,  $d_{\mathcal{Y}}$ , and  $d_{\mathcal{Z}}$  by solving a system of linear equations. In what follows, we will show that this can be further simplified, taking advantage of the special structure of the problem.

Consider the search direction defined by dual scaling (4.5). For  $i = 1, \dots, m$ , it can be decomposed as

$$\begin{aligned} & \begin{pmatrix} d_{\tau_i} \\ 0_d \\ d_{x_i} \end{pmatrix} + \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial z_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial s_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial z_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial s_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial s_i} \end{pmatrix} \begin{pmatrix} d_{t_i} \\ d_{z_i} \\ d_{s_i} \end{pmatrix} \\ &= -\frac{\rho}{\mathcal{X}^T \mathcal{S}} \begin{pmatrix} \tau_i \\ 0 \\ x_i \end{pmatrix} - \begin{pmatrix} \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial t_i} \\ \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial z_i} \\ \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial s_i} \end{pmatrix}. \end{aligned}$$

Note that  $s_i = c_i - A_i^T u$ ,  $d_{s_i} = -A_i^T d_u$ ,  $\tau_i = 1$ , and  $d_{\tau_i} = 0$  for  $i = 1, \dots, m$ . The system can be written as

$$\begin{aligned} \begin{pmatrix} 0 \\ 0 \\ d_{x_i} \end{pmatrix} + \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial z_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial s_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial z_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial s_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial s_i} \end{pmatrix} \begin{pmatrix} d_{t_i} \\ d_{z_i} \\ -A_i^T d_u \end{pmatrix} \\ = -\frac{\rho}{\mathcal{X}^T \mathcal{S}} \begin{pmatrix} 1 \\ 0 \\ x_i \end{pmatrix} - \begin{pmatrix} \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial t_i} \\ \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial z_i} \\ \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial s_i} \end{pmatrix}. \end{aligned}$$

Let

$$J_i = \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial z_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial t_i} & \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial z_i} \end{pmatrix}.$$

One can easily verify that  $J_i$  is positive definite. Therefore, we can compute  $J_i^{-1}$  in at most  $O(d^3)$  time. From the first two equations, we get

$$(5.1) \quad \begin{pmatrix} d_{t_i} \\ d_{z_i} \end{pmatrix} = J_i^{-1} \left[ \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial t_i \partial s_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial z_i \partial s_i} \end{pmatrix} A_i^T d_u - \frac{\rho}{\mathcal{X}^T \mathcal{S}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial t_i} \\ \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial z_i} \end{pmatrix} \right].$$

Substituting this into the third equation, we get

$$\begin{aligned} d_{x_i} + \left[ \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} \end{pmatrix}^T J_i^{-1} \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} \end{pmatrix} - \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial s_i} \right] A_i^T d_u \\ = \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} \end{pmatrix}^T J_i^{-1} \left[ \frac{\rho}{\mathcal{X}^T \mathcal{S}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial t_i} \\ \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial z_i} \end{pmatrix} \right] - \frac{\rho}{\mathcal{X}^T \mathcal{S}} x_i - \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial s_i}. \end{aligned}$$

Moreover, since

$$\sum_{i=1}^m A_i x_i = 0, \quad \sum_{i=1}^m A_i d_{x_i} = 0,$$

we have

$$\begin{aligned}
 (5.2) \quad & \sum_{i=1}^m A_i \left[ \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} \end{pmatrix}^T J_i^{-1} \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} \end{pmatrix} - \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial s_i} \right] A_i^T d_u \\
 & = \sum_{i=1}^m A_i \left( \begin{pmatrix} \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial t_i} \\ \frac{\partial^2 \hat{f}_p(t_i, s_i, z_i)}{\partial s_i \partial z_i} \end{pmatrix}^T J_i^{-1} \right. \\
 & \quad \left. \times \left[ \frac{\rho}{\mathcal{X}^T \mathcal{S}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial t_i} \\ \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial z_i} \end{pmatrix} \right] - \frac{\rho}{\mathcal{X}^T \mathcal{S}} x_i - \frac{\partial \hat{f}_p(t_i, s_i, z_i)}{\partial s_i} \right).
 \end{aligned}$$

Note that the system for computing  $d_u$  may not have full rank. If that is the case, any feasible solution is acceptable.

It requires  $O(m(d^3 + nd^2 + n^2d))$  operations to set up the system (5.2) for computing  $d_u$ . Solving the system requires  $O(n^3)$  operations. Once  $d_u$  is computed,  $O(m(d^3 + nd^2 + n^2d))$  operations are required to compute  $d_{\mathcal{X}}$ ,  $d_{\mathcal{S}}$ , and  $d_{\mathcal{Z}}$ . Therefore, the number of arithmetic operations in each iteration is bounded by  $O(n^3 + md^3 + md^2n + mdn^2)$ . The following theorem follows from Corollary 5.1 and the above analysis.

**THEOREM 5.3.** *Let the initial feasible primal solution  $\mathcal{X}^0$  and dual feasible solution  $(\mathcal{Y}^0; \mathcal{S}^0)$  be given as above. Then, an  $\epsilon$ -optimal solution to problem (2.1) can be obtained by the potential reduction algorithms in at most*

$$O\left(\gamma\sqrt{200(1+2d)m}(\log(\bar{c}/\epsilon) + \log m + \log d)\right)$$

iterations, where

$$\bar{c} = \max_{1 \leq i \leq m} \|c_i\|,$$

and each iteration requires  $O(n^3 + md^3 + md^2n + mdn^2)$  arithmetic operations.  $\square$

Note that if both  $d$  and  $\gamma$  are constants and the problem is normalized such that  $\bar{c} = 1$ , i.e., all of  $c_i$  is within the unit ball in  $R^d$ , then the iteration bound is  $O(\sqrt{m}(\log(1/\epsilon) + \log m))$ . We will further discuss this issue in the following applications.

**6. Applications.** In this section, we will apply the algorithms presented in the previous sections to solve the  $p$ -norm multifacility location (PMFL) problem and the  $p$ -norm SMT problem under a given topology. We will also take advantage of the special structures of these special problems and obtain improved computational complexity results wherever possible.

**6.1. The  $p$ -norm multifacility location problem.** Let  $a_1, a_2, \dots, a_M$  be  $M$  points in  $R^d$ , the  $d$ -dimensional  $l_p$  space. Let  $w_{ji}, j = 1, 2, \dots, N, i = 1, 2, \dots, M$ , and  $v_{jk}, 1 \leq j < k \leq N$ , be given nonnegative numbers. Find a point  $x = (x_1; x_2; \dots; x_N) \in R^{dN}$  that will minimize

$$(6.1) \quad f_p(x) = \sum_{j=1}^N \sum_{i=1}^M w_{ji} \|x_j - a_i\|_p + \sum_{1 \leq j < k \leq N} v_{jk} \|x_j - x_k\|_p, \quad p \geq 1.$$

This is the so-called PMFL. For ease of notation, we assume that  $v_{jj} = 0$  for  $j = 1, 2, \dots, N$  and that  $v_{jk} = v_{kj}$  for  $1 \leq k < j \leq N$ .

In the PMFL problem,  $a_1, a_2, \dots, a_M$  represent the locations of  $M$  existing facilities;  $x_1, x_2, \dots, x_N$  represent the locations of  $N$  new facilities; and the objective function  $f_p(x)$  is the sum of weighted  $p$ -norm distances from each new facility to each existing facility and those between each pair of new facilities. Our goal is to find optimal locations for the new facilities, i.e., to minimize  $f_p(x)$ .

In problem (6.1), some of the weights  $w_{ji}$  and  $v_{jk}$  may be zero. Let  $m$  be the number of nonzero weights in (6.1). Then the PMFL problem (6.1) is the minimization of a sum of  $m$   $p$ -norms. Without loss of generality, we assume that for each  $j \in \{1, 2, \dots, N\}$ , there exists a nonzero  $w_{ji}$  for some  $i \in \{1, 2, \dots, M\}$  or a nonzero  $v_{jk}$  for some  $k \in \{1, 2, \dots, N\}$ .

To transform the PMFL problem (6.1) into an instance of problem (1.1), we simply do the following. Let  $u = (x_1; x_2; \dots; x_N)$ . It is clear that  $u \in R^n$  where  $n = dN$ . For each nonzero  $w_{ji}$ , there is a corresponding term of  $p$ -norm  $\|c(w_{ji}) - A(w_{ji})^T u\|_p$  where  $c(w_{ji}) = w_{ji}a_i$  and  $A(w_{ji})^T$  is a row of  $N$  blocks of  $d$  by  $d$  matrices whose  $j$ th block is  $w_{ji}I_d$  and whose other blocks are all zero. For each nonzero  $v_{jk}$ , there is a corresponding term of  $p$ -norm  $\|c(v_{jk}) - A(v_{jk})^T u\|_p$ , where  $c(v_{jk}) = 0$ , and  $A(v_{jk})^T$  is a row of  $N$  blocks of  $d$  by  $d$  matrices whose  $j$ th and  $k$ th blocks are  $-v_{jk}I_d$  and  $v_{jk}I_d$ , respectively, and whose other blocks are all zero.

Now it is clear that we have transformed the PMFL problem (6.1) into an instance of (2.1), where  $n = dN$ , and  $m$  is the number of nonzero weights  $w_{ji}$  and  $v_{jk}$ . Note that the system (5.2) can be set up with  $O(md^2)$  operations. Therefore, it follows from Theorem 5.1 that we have Theorem 6.1.

**THEOREM 6.1.** *An  $\epsilon$ -optimal solution to the PMFL problem (6.1) can be computed using any of our algorithms in at most*

$$O\left(\gamma\sqrt{200(1+2d)MN}(\log(\bar{c}/\epsilon) + \log(MN) + \log d)\right)$$

*iterations, where  $\bar{c} = \max_{1 \leq j \leq n} \max_{1 \leq i \leq m} \|w_{ji}a_i\|$ , and each iteration requires  $O(d^3N^3 + MNd^2)$  arithmetic operations.  $\square$*

**6.2. The  $p$ -norm SMT problem.** The  $p$ -norm SMT problem is given by a set of points  $P = \{p_1, p_2, \dots, p_N\}$  in the  $l_p$ -plane and asks for the shortest planar straight-line graph spanning  $P$ . The solution takes the form of a tree, called the *Steiner minimal tree*, 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 that the degree of each Steiner point is at most 3. See [13, 22] for details.

**DEFINITION 6.2** (see [13, 15, 16]). *A full Steiner topology of point set  $P$  is a tree graph whose vertex set contains  $P$  and  $N - 2$  Steiner points; the degree of each vertex in  $P$  is exactly 1, and the degree of each Steiner vertex is exactly 3.*

Computing a SMT for a given set of  $N$  points in the  $l_p$ -plane is NP-hard [12]. However, the problem of computing the shortest network under a given full Steiner topology can be solved efficiently. Recently, there have been increased interests in this latter problem and several algorithms have been proposed [15, 16, 31]. We will formulate this problem as a special case of problem (1.1).

Let  $m = 2N - 3$ ,  $d = 2$ ,  $n = 2N - 4$ . Let  $u \in R^{2N-4}$  represent the locations of the  $N - 2$  Steiner points. Without loss of generality, we may order the edges in the given full Steiner topology in such a way that each of the first  $N$  edges connects a regular point to a Steiner point. For  $i = 1, 2, \dots, N$ ,  $c_i$  is  $p_{i_1}$ , where  $i_1$  is the index of

TABLE 1  
The coordinates of the 10 regular points.

Index	$x$ -coordinate	$y$ -coordinate	Index	$x$ -coordinate	$y$ -coordinate
9	2.30946900	9.20821100	14	7.59815200	0.61583600
10	0.57736700	6.48093800	15	8.56812900	3.07917900
11	0.80831400	3.51906200	16	4.75750600	3.75366600
12	1.68591200	1.23167200	17	3.92609700	7.00879800
13	4.11085500	0.82111400	18	7.43649000	7.68328400

TABLE 2  
The tree topology.

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			

the regular point on the  $i$ th edge;  $A_i^T \in R^{2 \times n}$  is a row of  $N - 2$  2 by 2 block matrices, where only the  $i_2$ th block is  $I_2$  and the rest are all zero, where  $i_2$  is the index of the Steiner point on the  $i$ th edge. For  $i = N + 1, N + 2, \dots, m$ ,  $c_i = 0$ , and  $A_i^T \in R^{2 \times n}$  is a row of  $N - 2$  2 by 2 block matrices, where the  $i_1$ st block is  $-I_2$ , the  $i_2$ nd block is  $I_2$ , and the rest of the blocks are all zero, and where  $i_1$  and  $i_2$  are the indices of the two Steiner points on the  $i$ th edge. It is clear that we have transformed the problem of computing a shortest network under a full Steiner topology into an instance of (2.1), where  $d = 2$ ,  $n = 2N - 4$ , and  $m = 2N - 3$ . Therefore, it can be solved efficiently using our interior point algorithm.

Note that we can move the point set  $P$  so that its gravitational center is the origin. Therefore, the  $l_p$ -norms (as well as the Euclidean norms) of the regular points are bounded by the largest pairwise ( $p$ -norm or Euclidean) distance among the points in  $P$ , which corresponds to the constant  $\bar{c}$  in previous theorems. Furthermore, as illustrated in [36], the search direction can be computed in  $O(N)$  arithmetic operations using a technique known as *Gaussian elimination on leaves of a tree* [31]. Therefore, we have the following theorem.

**THEOREM 6.3.** *An  $\epsilon$ -optimal solution to the shortest network under a given full Steiner topology of  $N$  regular points in the  $l_p$ -plane can be computed using our potential reduction algorithms in at most  $O(\sqrt{N}(\log(\bar{c}/\epsilon) + \log N))$  iterations, where  $\bar{c}$  is the largest pairwise distance among the regular points and each iteration requires  $O(N)$  arithmetic operations. Therefore, the computation of an  $\epsilon$ -optimal solution requires  $O(N\sqrt{N}(\log(\bar{c}/\epsilon) + \log N))$  arithmetic operations.  $\square$*

The problem of computing the shortest network under a full Steiner topology was first studied by Hwang [15], Hwang and Weng [16], and Smith [31]. Hwang [15] presented a linear-time exact algorithm that can output the shortest network under a given full Steiner topology if there exists a nondegenerate SMT corresponding to that given topology and that quits otherwise. Hwang and Weng [16] presented an  $O(N^2)$ -time graphical algorithm that can output the shortest network under a

TABLE 3  
*Output for  $p = 1.01$  using dual scaling.*

Iteration	Network-cost	Duality-gap	dstep-max	pstep-max
1	1.9810e+03	1.9863e+03	3.1250e-01	9.7656e-03
2	9.3134e+02	9.3948e+02	3.1250e-01	9.7656e-03
3	4.6604e+02	4.6352e+02	3.1250e-01	7.8125e-02
4	2.6185e+02	2.5141e+02	3.1250e-01	1.9531e-02
5	1.6673e+02	1.5389e+02	3.1250e-01	4.8828e-03
6	1.1164e+02	1.0050e+02	3.1250e-01	1.9531e-02
7	8.2320e+01	7.0048e+01	3.1250e-01	4.8828e-03
8	6.3980e+01	4.8173e+01	3.1250e-01	4.8828e-03
9	5.1500e+01	3.2779e+01	3.1250e-01	4.8828e-03
10	4.3166e+01	2.2726e+01	3.1250e-01	2.4414e-03
11	3.7916e+01	1.5499e+01	3.1250e-01	1.2207e-03
12	3.4452e+01	1.0726e+01	3.1250e-01	6.1035e-04
13	3.2373e+01	6.5718e+00	3.1250e-01	6.1035e-04
14	3.0782e+01	3.5208e+00	3.1250e-01	6.1035e-04
15	2.9691e+01	2.0491e+00	3.1250e-01	3.0518e-04
16	2.9255e+01	1.3093e+00	3.1250e-01	7.6294e-05
17	2.9023e+01	6.9142e-01	3.1250e-01	7.6294e-05
18	2.8824e+01	3.9670e-01	3.1250e-01	3.8147e-05
19	2.8737e+01	2.3084e-01	3.1250e-01	1.9073e-05
20	2.8682e+01	1.4074e-01	3.1250e-01	9.5367e-06
21	2.8649e+01	9.0313e-02	3.1250e-01	4.7684e-06
22	2.8630e+01	5.2046e-02	3.1250e-01	4.7684e-06
23	2.8615e+01	2.7875e-02	3.1250e-01	4.7684e-06
24	2.8605e+01	1.7067e-02	3.1250e-01	2.3842e-06
25	2.8601e+01	1.1235e-02	3.1250e-01	5.9605e-07
26	2.8599e+01	6.6474e-03	3.1250e-01	5.9605e-07
27	2.8597e+01	3.5436e-03	3.1250e-01	5.9605e-07
28	2.8596e+01	2.0801e-03	3.1250e-01	5.9605e-07
29	2.8595e+01	1.3643e-03	3.1250e-01	7.4506e-08
30	2.8594e+01	5.2467e-04	6.2500e-01	7.4506e-08
31	2.8594e+01	2.8420e-04	6.2500e-01	1.8626e-08
32	2.8594e+01	1.5954e-04	6.2500e-01	9.3132e-09
33	2.8594e+01	9.4942e-05	6.2500e-01	4.6566e-09
34	2.8594e+01	4.2928e-05	6.2500e-01	4.6566e-09
35	2.8594e+01	1.6635e-05	6.2500e-01	2.3283e-09
36	2.8594e+01	9.0572e-06	6.2500e-01	5.8208e-10

given full Steiner topology if the shortest network under the given topology is a tree with maximum vertex degree 3 and that quits otherwise. Xue and Ye [36] proposed a primal-dual potential reduction algorithm that can always output an  $\epsilon$ -optimal network under a given topology in  $O(N\sqrt{N}(\log(\bar{c}/\epsilon) + \log N))$  operations, where  $\bar{c}$  is the largest pairwise distance among the given points. It seems hard to generalize the graphical methods to the  $p$ -norm case. Therefore, our generalization of the result of [36] is important.

**7. Computational examples.** We have implemented all three versions of our algorithm using Matlab. Our intention here is to justify the theory developed in this paper. Therefore, our primary interest is in the number of iterations required by the algorithms. Our implementation here is very preliminary. Extensive computational study of the algorithms will be reported in a separate paper. For test problems, we have taken the 10-regular-points SMT problem from [36]. The coordinates of the regular points are given in Table 1. The tree topology is given in Table 2, where for each edge, indices of its two vertices are shown next to the index of the edge.

In our implementation, we used  $\gamma = 2m$  to take *long steps* instead of using the



TABLE 4  
Output for  $p = 1.50$  using dual scaling.

Iteration	Network-cost	Duality-gap	dstep-max	pstep-max
1	6.3022e+02	6.3501e+02	3.1250e-01	9.7656e-03
2	3.0949e+02	3.0829e+02	3.1250e-01	7.8125e-02
3	2.4246e+02	2.3339e+02	1.5625e-01	9.7656e-03
4	1.4290e+02	1.2897e+02	3.1250e-01	9.7656e-03
5	9.0251e+01	7.7152e+01	3.1250e-01	1.9531e-02
...				
23	2.6561e+01	6.2128e-03	3.1250e-01	5.9605e-07
24	2.6559e+01	3.7579e-03	3.1250e-01	2.9802e-07
25	2.6558e+01	2.3707e-03	3.1250e-01	1.4901e-07
26	2.6557e+01	7.8395e-04	6.2500e-01	1.4901e-07
27	2.6557e+01	3.5326e-04	6.2500e-01	3.7253e-08
28	2.6557e+01	1.4104e-04	6.2500e-01	1.8626e-08
29	2.6557e+01	7.4709e-05	6.2500e-01	4.6566e-09
30	2.6557e+01	4.4346e-05	6.2500e-01	2.3283e-09
31	2.6557e+01	1.8446e-05	6.2500e-01	2.3283e-09
32	2.6557e+01	9.9644e-06	3.1250e-01	1.1642e-09

TABLE 5  
Output for  $p = 2.00$  using dual scaling.

Iteration	Network-cost	Duality-gap	dstep-max	pstep-max
1	3.5099e+02	3.5513e+02	3.1250e-01	9.7656e-03
2	2.6731e+02	2.6493e+02	1.5625e-01	9.7656e-03
3	2.0753e+02	1.9761e+02	1.5625e-01	9.7656e-03
4	1.1948e+02	1.0726e+02	3.1250e-01	4.8828e-03
5	7.5355e+01	6.1541e+01	3.1250e-01	4.8828e-03
...				
28	2.5357e+01	1.2411e-03	3.1250e-01	7.4506e-08
29	2.5357e+01	7.0021e-04	3.1250e-01	7.4506e-08
30	2.5356e+01	3.7630e-04	3.1250e-01	7.4506e-08
31	2.5356e+01	2.2183e-04	3.1250e-01	3.7253e-08
32	2.5356e+01	1.4140e-04	3.1250e-01	9.3132e-09
33	2.5356e+01	7.8081e-05	3.1250e-01	9.3132e-09
34	2.5356e+01	4.8133e-05	3.1250e-01	4.6566e-09
35	2.5356e+01	2.9645e-05	3.1250e-01	2.3283e-09
36	2.5356e+01	1.5629e-05	3.1250e-01	2.3283e-09
37	2.5356e+01	8.3688e-06	3.1250e-01	4.6566e-09

conservative theoretical parameter  $\gamma = 1$ . Also, we used 0.75 times the largest feasible step-size as the actual step-size rather than using the theoretical step-size or a line-search. The algorithm stops whenever the absolute duality gap is smaller than  $10^{-5}$ .

To test the flexibility of our algorithm, we have used the values of  $p = 1.01, 1.5, 2.0, 3.0,$  and  $101$  on the 10-point case. Although with a larger parameter, the algorithm based on barrier function 2 works better in our examples. All five cases were solved within 30 to 40 iterations. The output of the algorithm is presented in Tables 3–7. Note that for the last case, we have used the primal scaling algorithm, where  $q = 1.01$ . The second column in the tables shows the cost of the current network (i.e., the sum of  $p$ -norms in the current network). The third column shows the duality gap, which is an upper bound of the error in the cost of the current network to the cost of the optimal (shortest) network. The last two columns show the largest dual and primal feasible step-sizes,  $\bar{\beta}$  and  $\bar{\alpha}$ ; see the discussion at the end of section 4.

The shortest networks for the cases  $p = 1.01, 1.50, 3.0, 101$  are plotted in Figure 1, where regular points are labeled by + and Steiner points are labeled by o. The case

TABLE 6  
*Output for  $p = 3.00$  using dual scaling.*

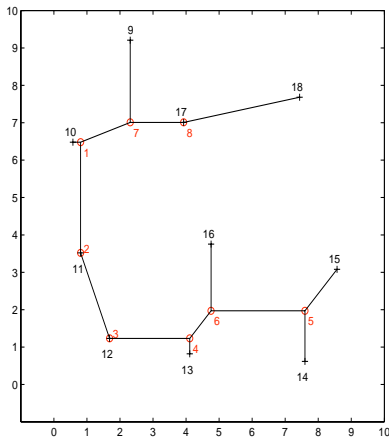
Iteration	Network-cost	Duality-gap	dstep-max	pstep-max
1	3.0811e+02	3.1301e+02	1.5625e-01	1.9531e-02
2	2.3295e+02	2.3135e+02	1.5625e-01	9.7656e-03
3	1.7991e+02	1.7375e+02	1.5625e-01	4.8828e-03
4	1.4208e+02	1.3316e+02	1.5625e-01	2.4414e-03
5	8.4541e+01	7.0951e+01	3.1250e-01	4.8828e-03
...				
24	2.3930e+01	5.6315e-03	3.1250e-01	5.9605e-07
25	2.3928e+01	2.9602e-03	3.1250e-01	5.9605e-07
26	2.3927e+01	1.6010e-03	3.1250e-01	5.9605e-07
27	2.3927e+01	1.0595e-03	3.1250e-01	3.7253e-08
28	2.3926e+01	4.9997e-04	6.2500e-01	3.7253e-08
29	2.3926e+01	2.0940e-04	1.2500e+00	1.8626e-08
30	2.3926e+01	6.3721e-05	2.5000e+00	9.3132e-09
31	2.3926e+01	2.7372e-05	2.5000e+00	2.3283e-09
32	2.3926e+01	1.2168e-05	1.2500e+00	1.1642e-09
33	2.3926e+01	5.4989e-06	6.2500e-01	5.8208e-10

TABLE 7  
*Output for  $p = 101$  ( $q = 1.01$ ) using primal scaling.*

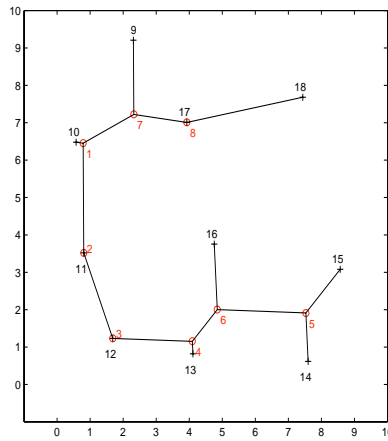
Iteration	Network-cost	Duality-gap	dstep-max	pstep-max
1	9.8742e+01	9.4987e+01	4.8828e-03	6.2500e-01
2	6.9168e+01	5.7033e+01	2.4414e-03	1.2500e+00
3	3.9986e+01	2.4935e+01	2.4414e-03	1.2500e+00
4	2.7451e+01	1.0022e+01	1.2207e-03	6.2500e-01
5	2.5071e+01	6.3698e+00	3.0518e-04	3.1250e-01
...				
23	2.1182e+01	1.3274e-04	9.3132e-09	3.1250e-01
24	2.1182e+01	5.1656e-05	9.3132e-09	3.1250e-01
25	2.1182e+01	2.6681e-05	9.3132e-09	1.5625e-01
26	2.1183e+01	2.7794e-04	1.4901e-07	1.5625e-01
27	2.1183e+01	1.6209e-04	9.3132e-09	5.0000e+00
28	2.1182e+01	9.0529e-05	4.6566e-09	1.2500e+00
29	2.1182e+01	5.8100e-05	2.3283e-09	2.5000e+00
30	2.1182e+01	2.7823e-05	2.3283e-09	1.2500e+00
31	2.1182e+01	1.3559e-05	1.1642e-09	6.2500e-01
32	2.1182e+01	6.2516e-06	5.8208e-10	6.2500e-01

$p = 2.0$  is similar but we choose not to illustrate it here to save space. It is interesting to see the slight change in the network when  $p$  is changed from 1.01 to 1.5 and eventually to 101.

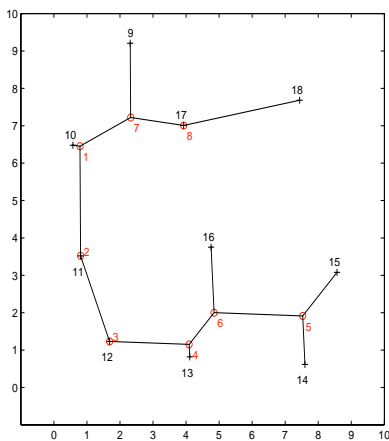
**8. Conclusions.** In this paper, we have transformed the problem of minimizing a sum of  $p$ -norms into a standard convex programming problem in conic forms. Unlike those in most convex optimization problems, the cone for this problem is not self-dual. We have constructed two barrier functions and studied its associated parameters. Using these barrier functions, we have presented a polynomial time primal-dual potential reduction algorithm for solving this problem. In particular, the number of iterations required to produce an  $\epsilon$ -optimal solution is at most  $O(\sqrt{md}(\log(\bar{c}/\epsilon) + \log(md)))$ . As applications, we have shown that computing an  $\epsilon$ -optimal solution of the shortest  $p$ -norm network under a tree topology interconnecting  $N$  regular points on the  $l_p$ -plane requires only  $O(N^{1.5}(\log(\bar{c}/\epsilon) + \log N))$  arithmetic operations, where  $\bar{c}$  is the largest pairwise  $l_p$ -distance among the given point set. Our implementation is only preliminary. Computational issues of our algorithm are under investigation and will be reported in another paper.



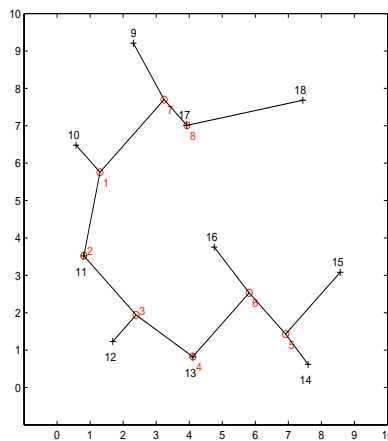
(a) The shortest network for  $p = 1.01$ .



(b) The shortest network for  $p = 1.50$ .



(c) The shortest network for  $p = 3.00$ .



(d) The shortest network for  $p = 101$ .

FIG. 1. Shortest networks for different values of  $p$ .

**Acknowledgments.** The authors would like to thank two referees and the associate editor for their helpful comments and remarks on the first version of the paper.

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, *A Symmetric Primal-Dual Newton Method for Minimizing a Sum of Norms*, manuscript, Odense University, Denmark, 1995.  
 [3] P.H. CALAMAI AND A.R. CONN, *A second-order method for solving the continuous multifacility location problem*, in Numerical Analysis: Proceedings of the Ninth Biennial Conference, Dundee, Scotland, Lecture Notes in Mathematics 912, G.A. Watson, ed., Springer-

- Verlag, New York, 1982, pp. 1–25.
- [4] P.H. CALAMAI AND A.R. CONN, *A projected Newton method for  $l_p$  norm location problems*, Math. Programming, 38 (1987), pp. 75–109.
  - [5] R. CHANDRASEKARAN AND A. TAMIR, *Open questions concerning Weiszfeld's algorithm for the Fermat-Weber location problem*, Math. Programming, 44 (1989), pp. 293–295.
  - [6] R. CHANDRASEKARAN AND A. TAMIR, *Algebraic optimization: The Fermat-Weber location problem*, Math. Programming, 46 (1990), pp. 219–224.
  - [7] K.D. ANDERSON, 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., to appear.
  - [8] D. DEN HERTOG, *Interior Point Approach to Linear, Quadratic and Convex Programming*, Kluwer Academic Publishers, Norwell, MA, 1994.
  - [9] D. DEN HERTOG, F. JARRE, C. ROOS, AND T. TERLAKY, *A sufficient condition for self-concordance, with application to some classes of structured convex programming problems*, Math. Programming, 69 (1995), pp. 75–88.
  - [10] J.W. EYSTER, J.A. WHITE, AND W.W. WIERWILLE, *On solving multifacility location problems using a hyperboloid approximation procedure*, AIIE Transactions, 5 (1973), pp. 1–6.
  - [11] R.L. FRANCIS, L.F. MCGINNIS, JR., AND J.A. WHITE, *Facility Layout and Location: An Analytical Approach*, Prentice-Hall, Englewood Cliffs, NJ, 1991.
  - [12] M.R. GAREY, R.L. GRAHAM, AND D.S. JOHNSON, *The complexity of computing Steiner minimal trees*, SIAM J. Appl. Math., 32 (1977), pp. 835–859.
  - [13] E.N. GILBERT AND H.O. POLLAK, *Steiner minimal trees*, SIAM J. Appl. Math., 16 (1968), pp. 1–29.
  - [14] G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1983.
  - [15] F.K. HWANG, *A linear time algorithm for full Steiner trees*, Oper. Res. Lett., 4 (1986), pp. 235–237.
  - [16] F.K. HWANG AND J.F. WENG, *The shortest network under a given topology*, J. Algorithms, 13 (1992), pp. 468–488.
  - [17] I.N. KATZ, *Local convergence in Fermat's problem*, Math. Programming, 6 (1974), pp. 89–104.
  - [18] M. KOJIMA, S. SHINDOH, AND S. HARA, *Interior-Point Methods for the Monotone Semidefinite Linear Complementarity Problem in Symmetric Matrices*, Research Reports on Information Sciences, No. B-282, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Japan, 1994; revised April, 1995.
  - [19] H.W. KUHN, *A note on Fermat's problem*, Math. Programming, 4 (1973), pp. 98–107.
  - [20] Y. LI, *A Newton Acceleration of the Weiszfeld Algorithm for Minimizing the Sum of Euclidean Distances*, Technical report TR95-1552, Computer Science Department, Cornell University, Ithaca, NY, November 1995.
  - [21] R.F. LOVE, J.G. MORRIS, AND G.O. WESOLOWSKY, *Facilities Location: Models & Methods*, North-Holland, Amsterdam, 1988.
  - [22] Z.A. MELZAK, *On the problem of Steiner*, Canad. Math. Bull., 16 (1961), pp. 143–148.
  - [23] W. MIEHLE, *Link length minimization in networks*, Oper. Res., 6 (1958), pp. 232–243.
  - [24] A. NEMIROVSKII, *Polynomial time methods in convex programming*, in The Mathematics of Numerical Analysis, Lectures in Applied Mathematics 32, J. Renegar, M. Shub, and S. Smale, eds., AMS, Providence, RI, 1996.
  - [25] YU. NESTEROV AND A. NEMIROVSKII, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Stud. Appl. Math. 13, SIAM, Philadelphia, PA, 1994.
  - [26] YU. E. NESTEROV AND M. J. TODD, *Self-scaled barriers and interior-point methods for convex programming*, Math. Oper. Res. 22 (1997), pp. 1–42.
  - [27] L.M. OSTRESH, *The multifacility location problem: Applications and descent theorems*, J. Regional Science, 17 (1977), pp. 409–419.
  - [28] M.L. OVERTON, *A quadratically convergent method for minimizing a sum of Euclidean norms*, Math. Programming, 27 (1983), pp. 34–63.
  - [29] 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.
  - [30] 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.
  - [31] W.D. SMITH, *How to find Steiner minimal trees in Euclidean  $d$ -space*, Algorithmica, 7 (1992), pp. 137–177.
  - [32] C.Y. WANG ET AL., *On the convergence and rate of convergence of an iterative algorithm for the plant location problem*, Qufu Shiyun Xuebao, 2 (1975), pp. 14–25 (in Chinese).

- [33] E. WEISZFELD, *Sur le point par lequel la somme des distances de  $n$  points donnés est minimum*, Tohoku Math. J., 43 (1937), pp. 355–386.
- [34] G.L. XUE, *Algorithms for Computing Extreme Points of Convex Hulls and the Euclidean Facilities Location Problem*, Ph.D. thesis, Computer Science Department, University of Minnesota, Minneapolis, 1991.
- [35] G.L. XUE, J.B. ROSEN, AND P.M. PARDALOS, *A polynomial time dual algorithm for the Euclidean multifacility location problem*, in Proceedings of Second Conference on Integer Programming and Combinatorial Optimization, Pittsburgh, PA, 1992, pp. 227–236.
- [36] 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.