An Efficient Primal-Dual Interior-Point Method for Minimizing a Sum of Euclidean Norms

Knud D. Andersen* Edmund Christiansen† Andrew R. Conn‡
Michael L. Overton§

NYU Computer Science Dept Technical Report 769
August 17, 1998

Abstract

The problem of minimizing a sum of Euclidean norms dates from the 17th century and may be the earliest example of duality in the mathematical programming literature. This nonsmooth optimization problem arises in many different kinds of modern scientific applications. We derive a primal-dual interior-point algorithm for the problem, by applying Newton's method directly to a system of nonlinear equations characterizing primal and dual feasibility and a perturbed complementarity condition. The main work at each step consists of solving a system of linear equations (the Schur complement equations). This Schur complement matrix is not symmetric, unlike in linear programming. We incorporate a Mehrotra-type predictor-corrector scheme and present some experimental results comparing several variations of the algorithm, including, as one option, explicit symmetrization of the Schur complement with a skew corrector term. We also present results obtained from a code implemented to solve large sparse problems, using a symmetrized Schur complement. This has been applied to problems arising in plastic collapse analysis, with hundreds of thousands of variables and millions of nonzeros in the constraint matrix. The algorithm typically finds accurate solutions in less than 50 iterations and determines physically meaningful solutions previously unobtainable.

*Dash Associates, Leamington Spa, England. This work was supported by a Danish SNF Research Studentship.

†Department of Mathematics and Computer Science, Odense University, Denmark

‡IBM Thomas J. Watson Research Center, Yorktown Heights, NY. Supported in part by the U.S. Department of Defense Advanced Research Projects Agency.

§Courant Institute of Mathematical Sciences, New York University, New York, NY. Supported in part by IBM Thomas J. Watson Research Center, in part by the National Science Foundation, and in part by the U.S. Department of Energy.

Keywords: Sum of Norms, Nonsmooth Optimization, Duality, Newton Method, Primal-Dual, Interior-Point Method, Fermat Problem.

1 Introduction

A problem which arises in many applications is to minimize a sum of Euclidean vector norms, i.e.

$$D: \quad \min \left\{ \sum_{i=1}^{n} \|\boldsymbol{z}_i\| \ : \ \boldsymbol{y} \in \Re^m; \ \boldsymbol{z}_i \in \Re^d; \ \boldsymbol{A}_i^T \boldsymbol{y} + \boldsymbol{z}_i = \boldsymbol{c}_i, \ i = 1, \ldots, n \right\}$$

where $\boldsymbol{A}_i \in \Re^{m \times d}$, $\boldsymbol{c}_i \in \Re^d$, $i = 1, \ldots, n$, are given. In most applications $d = 2$ or $d = 3$ so that the terms in the sum are distances in a two or three dimensional Euclidean space. If $d = 1$, the problem D is equivalent to a linear program (LP). The minimization objective is convex but not differentiable at any point where some $\boldsymbol{z}_i = 0$.

The sum of distances problem, D, has a long and interesting history. The special case $d = m = 2$, $n = 3$, $\boldsymbol{A}_i = \boldsymbol{I}$, was studied by Fermat in the 17th century. This amounts to finding the point in $\Re^2$ which minimizes the sum of distances from it to three given points. In the early 19th century it was realized that this particular convex optimization problem has a natural dual maximization formulation. Kuhn [Kuh91] regards this as the first instance of duality in the mathematical programming literature. Further history is given in [Kuh67].

Duality theory for D is easily described using min-max theory. Let $\boldsymbol{x}_i \in \Re^d$, $i = 1, \ldots, n$. For consistency with standard notation for LP, we refer to $\boldsymbol{x}_i$ as the primal variables and $\boldsymbol{y}$, $\boldsymbol{z}_i$ as the dual variables, even though in our experience it is usually the dual problem D which explicitly arises in applications. We have

$$
\begin{aligned}
\min_{A_i^T y + z_i = c_i} \sum_{i=1}^n \|\boldsymbol{z}_i\| &= \min_{A_i^T y + z_i = c_i} \max_{\|x_i\| \le 1} \sum_{i=1}^n \boldsymbol{x}_i^T \boldsymbol{z}_i \\
&= \max_{\|x_i\| \le 1} \min_{A_i^T y + z_i = c_i} \sum_{i=1}^n \boldsymbol{x}_i^T \boldsymbol{z}_i \\
&= \max_{\|x_i\| \le 1} \min_{y} \left(\sum_{i=1}^n \boldsymbol{c}_i^T \boldsymbol{x}_i - \boldsymbol{y}^T \sum_{i=1}^n \boldsymbol{A}_i \boldsymbol{x}_i \right) \\
&= \max \left\{ \sum_{i=1}^n \boldsymbol{c}_i^T \boldsymbol{x}_i \;:\; \|\boldsymbol{x}_i\| \le 1;\; \sum_{i=1}^n \boldsymbol{A}_i \boldsymbol{x}_i = \boldsymbol{0} \right\}.
\end{aligned}
$$

The first equality follows from Cauchy-Schwartz, the second from min-max theory [Roc70, Cor. 37.3.2], the third trivially, and the fourth because if $\sum_{i=1}^n \boldsymbol{A}_i \boldsymbol{x}_i$ is not zero, the minimized value would be $-\infty$. Therefore, the dual of D is the primal problem

$$
P: \quad \max \left\{ \sum_{i=1}^n \boldsymbol{c}_i^T \boldsymbol{x}_i \;:\; \boldsymbol{x}_i \in \Re^d,\ \|\boldsymbol{x}_i\| \le 1,\ i = 1, \ldots, n;\ \sum_{i=1}^n \boldsymbol{A}_i \boldsymbol{x}_i = \boldsymbol{0} \right\}.
$$

This result is an easy generalization of the duality theory in [Kuh67], but may not have explicitly appeared in the literature until [And96b].

Although the duality theory has been known in its simplest form for nearly two centuries, it was not understood until relatively recently how to exploit duality in algorithms for minimizing D. Iteratively reweighted least squares (Weiszfeld's method [Wei37]) has long been used as a robust though slowly converging method to solve D. Another well-known approach is to replace the terms $\|\boldsymbol{z}_i\|$ in the objective by the differentiable quantity $\sqrt{\|\boldsymbol{z}_i\|^2 + \mu^2}$, where μ is a fixed positive number. This method is also robust but converges arbitrarily slowly as $\mu \to 0$. Neither of these algorithms use any aspect of duality. In both cases, the reason for the slow convergence is that, in most interesting applications, some of the norms in the objective D have zero as their optimal value.

Calamai and Conn[CC80, CC87] and Overton[Ove83] solved D using Newton methods combined with an active set approach to determine which norms $\|\boldsymbol{z}_i\|$ are zero at an optimal solution. These methods were the first to exploit the duality structure of the problem, as they explicitly compute both primal and dual solutions. However, Newton's method was derived in the $\boldsymbol{y}, \boldsymbol{z}$ space only, with the $\boldsymbol{x}$ variables computed by least-squares estimates. The methods of Calamai and Conn and Overton are quite efficient if not many norms $\|z_i\|$ are zero. However, if this number is large, the number of iterations is typically also large, because the active set of zero norms must be updated at every step.

Andersen [And96b] gave a method for solving D which is based on a primal interior-point method for LP. In this method the terms $\|z_i\|$ are replaced by $\sqrt{\|z_i\|^2 + \mu^2}$, but the quantity μ is treated as an extra variable, whose value is determined by duality estimates. Using this method, Andersen was the first to be able to solve D rapidly and accurately even when the number of variables is large and many norms $\|z_i\|$ are zero at a solution point. In [AC98] it was demonstrated how the linearly constrained problem can be reduced to the unconstrained case using an exact l_1 penalty function, while still preserving sparsity structure.

In this paper we present a primal-dual interior-point method for solving P and D. The basic algorithm is easy to motivate and implement. The number of iterations required is substantially fewer than for the primal interior-point method used by Andersen [And96b]. This is consistent with general experience with primal-dual versus primal interior-point methods for LP [Wri97].

The sum of norms problem is a special case of quadratically constrained quadratic programming (QCQP), also known as optimization over the quadratic cone. Nesterov and Todd [NT98a, NT98b] give a theoretical discussion of algorithms for optimization over homogeneous self-dual cones, including the quadratic cone. See also Adler and Alizadeh [AA95] for another primal-dual algorithmic approach to QCQP. Our view is that the sum of norms problem is sufficiently important that a specialized approach is justified. Also taking this view, Xue and Ye [XY97] give a complexity analysis of the sum of norms problem, using an interior-point method and exploiting the general theory given in [NT98a, NT98b].

Our primal-dual algorithm is derived in the next section, applying Newton's method to three conditions: primal and dual feasibility and complementarity. A key point is the derivation of the appropriate complementarity condition. The main work at each step consists of solving a system of linear equations (the Schur complement equations). This Schur complement matrix is not symmetric, unlike its counterpart in linear programming.

Section 3 discusses a Mehrotra predictor-corrector enhancement to the algorithm and considers symmetrizing the Schur complement equations, including a compensating skew corrector term. Section 4 presents experimental results for some Steiner tree test problems.

Section 5 discusses a large-scale implementation using a symmetrized Schur complement. This has been used to solve applied problems arising in plastic collapse analysis with hundreds of thousands of variables and millions of nonzeros in the constraint matrix. The algorithm typically finds accurate solutions in less than 50 iterations and determines physically meaningful solutions that were considered unobtainable until now.

In fact, problem D arises in many applications. Alpert, Chan, Kahng, Markov and Mulet [ACK$^+$98] have recently applied a variant of our algorithm presented in this paper to the placement of circuits in VLSI design. Chan, Golub and Mulet [CGM96] applied a nonlinear version of the algorithm to some applications in image reconstruction. Byrnes and Bright [BB] used iteratively reweighted least-squares to solve trajectory optimization problems in space exploration. In fact, this method (Weiszfeld's method) has long been used at the Jet Propulsion Laboratory as a basic workhorse to solve problems of the form D that arise in spacecraft missions such as the Galileo and Pioneer "fly-by's" of the outer planets. Strang [Str79] considered an isoparametric design problem to which Overton [Ove84] applied a version of his algorithm mentioned above. Parks [Par91] has applied related methods to solve minimal surface (soap bubble) problems. Alexander and Maddocks [AM93] used the method of [Ove83] to solve friction problems arising in robotics. A key similarity in all these applications is that some, and perhaps many, of the norms in the sum to be minimized can expected to have the value zero at an optimal solution.

We believe there is great opportunity to apply the primal-dual method given in this paper to these and many other interesting applications.

Notation. Let $\boldsymbol{I}_d$ denote the $d \times d$ identity matrix. Let

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} \in \Re^{dn}, \quad \boldsymbol{z} = \begin{bmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_n \end{bmatrix} \in \Re^{dn}, \quad \boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_n \end{bmatrix} \in \Re^{dn},$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \cdots & \boldsymbol{A}_n \end{bmatrix} \in \Re^{m \times dn}.$$

The primal feasible region is given by

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \Re^{dn} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}; \quad \|\boldsymbol{x}_i\| \leq 1,\ i = 1, \ldots, n \right\} \tag{1}$$

and the dual feasible region is

$$\mathcal{Y} = \left\{ (\boldsymbol{y}, \boldsymbol{z}) \in \Re^m \times \Re^{dn} : \boldsymbol{A}^T \boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c} \right\}. \tag{2}$$

Consequently, we may rewrite D as

$$D: \quad \min \left\{ \sum_{i=1}^{n} \|\boldsymbol{z}_i\| \ : \ (\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \right\}$$

and P as

$$P: \quad \max \ \left\{ \boldsymbol{c}^T \boldsymbol{x} \ : \ \boldsymbol{x} \in \mathcal{X} \right\}.$$

2 Complementarity and Newton's Method

Suppose $\boldsymbol{x} \in \mathcal{X}$ and $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y}$ are respectively primal and dual feasible. Then the duality gap, i.e. difference between the primal and dual objective functions, is

$$\sum_{i=1}^{n} \|\boldsymbol{z}_i\| - \sum_{i=1}^{n} \boldsymbol{c}_i^T \boldsymbol{x}_i = \sum_{i=1}^{n} \left(\|\boldsymbol{z}_i\| - \boldsymbol{x}_i^T \boldsymbol{z}_i \right) \geq 0. \tag{3}$$

The duality gap must be zero at an optimal solution. It is zero if and only if, for each $i = 1, \ldots, n$, either $\|\boldsymbol{z}_i\|$ is zero or $\boldsymbol{x}_i = \boldsymbol{z}_i / \|\boldsymbol{z}_i\|$. This complementarity condition can be conveniently expressed as

$$\boldsymbol{z}_i - \|\boldsymbol{z}_i\| \boldsymbol{x}_i = \boldsymbol{0}, \quad i = 1, \ldots, n. \tag{4}$$

It follows from the complementarity condition that for each i, either $\boldsymbol{z}_i = 0$ or $\|\boldsymbol{x}_i\| = 1$; we say that strict complementarity holds if, for each i, only one of these two conditions holds. It may happen that no strictly complementary solution exists, unlike in LP.

Primal-dual interior-point methods are based on Newton's method applied to three sets of equations: primal feasibility, dual feasibility, and an appropriate complementarity/centering condition. The feasibility equations are respectively

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \tag{5}$$

and

$$\boldsymbol{A}^T \boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c}. \tag{6}$$

We assume from now on that the m by dn matrix $\boldsymbol{A}$ has full rank. We also assume that $m < dn$, since otherwise P and D are solved by $\boldsymbol{x} = \boldsymbol{0}$, $\boldsymbol{z} = \boldsymbol{0}$.

The primal and dual feasibility equations consist of $m + dn$ equations in the $m + 2dn$ scalar variables represented by $\boldsymbol{y}$ and $\boldsymbol{x}$, $\boldsymbol{z}$. To make this a square system we need another dn equations, which are available in the form of the complementarity condition (4). This condition is not differentiable if $\|\boldsymbol{z}_i\|$ is zero, but it may be replaced by the centering condition

$$\boldsymbol{z}_i - \left(\|\boldsymbol{z}_i\|^2 + \mu^2\right)^{\frac{1}{2}} \boldsymbol{x}_i = \boldsymbol{0}, \quad i = 1, \ldots, n \tag{7}$$

where $\mu > 0$.

The following theorem is from [And96b], showing that the centering condition (7) is in fact the complementarity condition for the following pair of smooth optimization problems:

$$D_\mu : \quad \min \left\{ \sum_{i=1}^{n} \left(\|\boldsymbol{z}_i\|^2 + \mu^2\right)^{\frac{1}{2}} : \ (\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \right\}$$

$$P_\mu : \quad \max \left\{ \boldsymbol{c}^T \boldsymbol{x} + \mu \sum_{i=1}^{n} \left(1 - \|\boldsymbol{x}_i\|^2\right)^{\frac{1}{2}} : \ \boldsymbol{x} \in \mathcal{X} \right\}.$$

Theorem 1 *The problems D_μ and P_μ are a primal-dual pair. Specifically, D_μ has the solution $(\boldsymbol{y}(\mu), \boldsymbol{z}(\mu))$ and P_μ has the solution $\boldsymbol{x}(\mu)$, all satisfying (5), (6) and (7).*

Proof: The proof is a simple modification of the proof (given in Section 1) that P and D are a primal-dual pair. See [And96b] for details. □

This theorem shows that introducing the centering parameter μ in the complementarity conditions for the original pair of problems is equivalent to smoothing the norms in D and introducing a cost into P which moves the primal solution away from its boundary. The solutions $(\boldsymbol{x}(\mu), \boldsymbol{y}(\mu), \boldsymbol{z}(\mu))$ of P_μ, D_μ, for $\mu > 0$, define a sort of central path for P, D, though not one derived from a logarithmic barrier function and therefore not centered in the usual sense.

Let us write the centering condition (7) as

$$\Theta(\mu, \boldsymbol{z})\boldsymbol{x} - \boldsymbol{z} = \boldsymbol{0}, \quad \text{where} \quad \Theta(\mu, \boldsymbol{z}) = \mathrm{Diag}\left(\left(\|\boldsymbol{z}_i\|^2 + \mu^2\right)^{\frac{1}{2}} \boldsymbol{I}_d\right). \tag{8}$$

Collecting (6), (5) and (8) together, we have the nonlinear system of equations

$$\boldsymbol{G}_\mu(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \begin{pmatrix} \boldsymbol{A}^T \boldsymbol{y} + \boldsymbol{z} - \boldsymbol{c} \\ \boldsymbol{A}\boldsymbol{x} \\ \Theta(\mu, \boldsymbol{z})\boldsymbol{x} - \boldsymbol{z} \end{pmatrix} = \boldsymbol{0}, \tag{9}$$

whose solution is $(\boldsymbol{x}(\mu), \boldsymbol{y}(\mu), \boldsymbol{z}(\mu))$. Newton's method applied to $\boldsymbol{G}_\mu$ at a given point $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ gives the following linear system defining updates to the variables:

$$\begin{bmatrix} \mathbf{0} & \boldsymbol{A}^T & \boldsymbol{I}_{dn} \\ \boldsymbol{A} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{E}_\mu & \mathbf{0} & -\boldsymbol{F}_\mu \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{y} \\ \Delta \boldsymbol{z} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_d \\ \boldsymbol{r}_p \\ \boldsymbol{r}_c \end{bmatrix} \tag{10}$$

where

$$\boldsymbol{r}_d = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y} - \boldsymbol{z}, \quad \boldsymbol{r}_p = -\boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{r}_c = \boldsymbol{z} - \boldsymbol{E}_\mu \boldsymbol{x}, \tag{11}$$

$$\boldsymbol{E}_\mu = \operatorname{Diag}\left(\omega_i^\mu \boldsymbol{I}_d\right), \quad \boldsymbol{F}_\mu = \operatorname{Diag}\left(\boldsymbol{I}_d - \frac{1}{\omega_i^\mu}\boldsymbol{x}_i\boldsymbol{z}_i^T\right), \tag{12}$$

and

$$\omega_i^\mu = \left(\|\boldsymbol{z}_i\|^2 + \mu^2\right)^{\frac{1}{2}}. \tag{13}$$

The equation $\boldsymbol{r}_d = \mathbf{0}$ is maintained exactly at each step, by always defining $\boldsymbol{z} = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y}$. Although $\boldsymbol{r}_p$ can be set to zero initially by setting the first iterate $\boldsymbol{x} = \mathbf{0}$ or setting $\boldsymbol{x}$ to a null vector of $\boldsymbol{A}$, we do not assume this, since primal feasibility cannot be maintained exactly in the presence of rounding errors.

Eliminating $\Delta\boldsymbol{z}$, we have

$$\begin{bmatrix} \boldsymbol{F}_\mu^{-1}\boldsymbol{E}_\mu & \boldsymbol{A}^T \\ \boldsymbol{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}_\mu^{-1}\boldsymbol{r}_c \\ -\boldsymbol{A}\boldsymbol{x} \end{bmatrix} \tag{14}$$

Defining $\boldsymbol{H}_\mu = \boldsymbol{E}_\mu^{-1}\boldsymbol{F}_\mu$, and using the definition of $\boldsymbol{r}_c$ in (11), we find after one more elimination step that

$$\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T\Delta\boldsymbol{y} = \boldsymbol{A}\boldsymbol{E}_\mu^{-1}\boldsymbol{z} \tag{15}$$

and

$$\Delta\boldsymbol{x} = \boldsymbol{E}_\mu^{-1}\left(\boldsymbol{F}_\mu\Delta\boldsymbol{z} + \boldsymbol{r}_c\right), \tag{16}$$

where (immediately from dual feasibility)

$$\Delta\boldsymbol{z} = -\boldsymbol{A}^T\Delta\boldsymbol{y}. \tag{17}$$

The operations of multiplying vectors by $\boldsymbol{F}_\mu$ and $\boldsymbol{E}_\mu^{-1}$ are trivial since $\boldsymbol{F}_\mu$ is block diagonal and $\boldsymbol{E}_\mu$ is positive diagonal. Notice the explicit dependence of $\boldsymbol{E}_\mu$ and $\boldsymbol{F}_\mu$ on the centering parameter μ, in contrast to the situation in LP, where the corresponding diagonal matrices depend only on the current variables. This is a consequence of the more complicated nature of the complementarity condition (4).

The main cost of this process is forming and factoring the Schur complement $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$. Except in the trivial case $d = 1$, the block diagonal matrix $\boldsymbol{H}_\mu$ is not generally symmetric, since $\boldsymbol{F}_\mu$ is not. This presents difficulties for large sparse problems, where it is highly desirable to use sparse Cholesky techniques which apply only to symmetric, positive definite systems. Sparse LU factorization techniques for nonsymmetric linear systems are more costly since they require pivoting for stability and are therefore not able to exploit sparsity as effectively as sparse Cholesky methods.

However, note that $\boldsymbol{F}_\mu$ is positive definite (in the sense that $v^T\boldsymbol{F}_\mu v > 0$ for all $v \neq 0$) as long as $\boldsymbol{x} \in \mathcal{X}$, and therefore so are $\boldsymbol{H}_\mu$ and $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$ (since $\boldsymbol{E}_\mu$ and $\boldsymbol{F}_\mu$ commute). Furthermore, it follows from equation (7) that for $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = (\boldsymbol{x}(\mu), \boldsymbol{y}(\mu), \boldsymbol{z}(\mu))$ (see Theorem 1), the matrix $\boldsymbol{F}_\mu$ and therefore also $\boldsymbol{H}_\mu$ and $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$ are symmetric for all $\mu > 0$. Consequently, when defined sufficiently close to the "central path", $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$ is nearly symmetric. One of the issues we shall discuss in the next section is the effect of symmetrizing $\boldsymbol{H}_\mu$ by defining it to be $\frac{1}{2}\boldsymbol{E}_\mu^{-1}(\boldsymbol{F}_\mu + \boldsymbol{F}_\mu^T)$ instead of $\boldsymbol{E}_\mu^{-1}\boldsymbol{F}_\mu$.

As $\mu \to 0$, for each i, either $\boldsymbol{z}_i(\mu)$ or $\boldsymbol{x}_i(\mu) - \boldsymbol{z}_i(\mu)/\|\boldsymbol{z}_i(\mu)\|$ converges to zero. In the the latter case, the limit of the ith block of the corresponding $\boldsymbol{F}_\mu$ is singular, while in the former case the limit of the ith block of the corresponding $\boldsymbol{E}_\mu$ is zero.

We now discuss how to update the iterates $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ after $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{y}$ and $\Delta\boldsymbol{z}$ are computed. We start by observing that $\Delta\boldsymbol{z}$ is a descent direction for the smoothed dual objective function in D_μ: the gradient of this function with respect to $\boldsymbol{z}$ is easily seen to be $\boldsymbol{E}_\mu^{-1}\boldsymbol{z}$. Using (17) and (15), we have

$$\begin{aligned}(\Delta\boldsymbol{z})^T\boldsymbol{E}_\mu^{-1}\boldsymbol{z} &= -(\Delta\boldsymbol{y})^T\boldsymbol{A}\boldsymbol{E}_\mu^{-1}\boldsymbol{z}\\ &= -(\Delta\boldsymbol{y})^T\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T\Delta\boldsymbol{y}\\ &< 0\end{aligned}$$

(unless $\boldsymbol{z} = \boldsymbol{0}$), since the symmetric part of $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$ is positive definite. Consequently, it is natural to update $\boldsymbol{y}$ and $\boldsymbol{z}$ by using a line search on the smoothed dual function in D_μ. We therefore update the dual variables by

Dual Line Search Rule

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} + \hat{\beta}\Delta\boldsymbol{y}, \quad \tilde{\boldsymbol{z}} = \boldsymbol{z} + \hat{\beta}\Delta\boldsymbol{z}, \tag{18}$$

where

$$\hat{\beta} \approx \arg \min_{0 \le \beta \le 1} \sum_{i=1}^{n} \left(\|\boldsymbol{z}_i + \beta \Delta \boldsymbol{z}_i\|^2 + \mu^2 \right)^{\frac{1}{2}} \tag{19}$$

The same steplength must be used for $\boldsymbol{y}$ and $\boldsymbol{z}$ to maintain dual feasibility. Of course, the univariate minimization problem need not be solved exactly.

The direction $\Delta \boldsymbol{x}$ is not necessarily an ascent direction for the penalized primal objective function in P_μ, so a line search is not appropriate to update the primal iterate $\boldsymbol{x}$. We consider two possibilities:

Primal Scaling Rule

$$\tilde{\boldsymbol{x}} = \bar{\gamma} \left(\boldsymbol{x} + \Delta \boldsymbol{x} \right) \tag{20}$$

where

$$\bar{\gamma} = \max \left\{ \gamma : \quad \gamma \|\boldsymbol{x}_i + \Delta \boldsymbol{x}_i\| \le 1 \quad i = 1, \ldots, n \right\}. \tag{21}$$

The scaling rule is a trivial computation.

Primal Steplength Rule The step to the boundary is given by

$$\alpha_{\max} = \max \left\{ \alpha : \ \|\boldsymbol{x}_i + \alpha \Delta \boldsymbol{x}_i\| \le 1, \ i = 1, \ldots, n \right\} \ = \ \max_i \alpha_i,$$

where α_i is the positive root of a quadratic equation:

$$\alpha_i = \frac{-\Delta \boldsymbol{x}_i^T \boldsymbol{x}_i + \sqrt{\|\Delta \boldsymbol{x}_i\|^2 \left(1 - \|\boldsymbol{x}_i\|^2\right) + \left(\Delta \boldsymbol{x}_i^T \boldsymbol{x}_i\right)^2}}{\|\Delta \boldsymbol{x}_i\|^2}, \quad i = 1, \ldots, n.$$

We choose $0 \ll \tau < 1$ and define

$$\bar{\alpha} = \tau \alpha_{\max}. \tag{22}$$

Then the steplength rule is defined by

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \min(1, \bar{\alpha}) \Delta \boldsymbol{x}. \tag{23}$$

Both rules preserve primal feasibility in exact arithmetic. For the steplength rule, conventional experience with primal-dual interior-point methods dictates a choice of τ less than 1, but not much less, for example, $\tau = .99$ or $\tau = .999$. For sums of norms, however, we found that $\tau = 1$ also works quite well. This allows iterates $\boldsymbol{x}$ to actually reach the boundary of the feasible region, but the matrix $\boldsymbol{F}_\mu$ still cannot be singular as long as $\mu > 0$. Increased ill-conditioning of the linear systems which are solved as convergence takes place is a standard feature of interior-point methods and generally does not cause great difficulties except when the iterates are nearly optimal. The reason we do not allow $\tau = 1$ is that rounding errors may then cause the updated $\boldsymbol{x}$ to lie just outside the feasible region.

The scaling rule always places $\boldsymbol{x}$ exactly on the boundary of the feasible region. This is not appropriate if the solution $\boldsymbol{x}$ has all component norms $\|\boldsymbol{x}_i\| < 1$, but this is a trivial case since then the dual solution must be zero by complementarity. As far as we know, the scaling rule does not have an analogy in standard interior-point implementations: such a rule is possible only when the primal equality constraints are homogeneous as they are here.

Equations (15), (16), (17) and the updating rules just described define the basic ingredients of a primal-dual interior-point method for solving P and D. To complete the description of the algorithm we must define a rule for updating the parameter μ: for this we introduce a predictor-corrector method.

3 Mehrotra's Predictor-Corrector Method and a Symmetrized Algorithm with a Skew Correction Term

Mehrotra's predictor-corrector method is a standard tool in primal-dual interior-point software for LP. The basic idea is to first compute a predictor step defined by first-order approximations to the optimality conditions (i.e. Newton's method), and to follow this with a corrector step which also takes second-order terms into account. A key point is that both predictor and corrector use the same matrix factorization; only the right-hand sides of the linear equations defining the steps differ. Another key component is a technique for estimating the centering parameter μ. Mehrotra's method was originally given in [Meh92]; an excellent discussion may be found in [Wri97, Chap. 10].

We now discuss how to adapt Mehrotra's method to our problem. Let us replace $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ in the centering condition (8) by $\boldsymbol{x} + \Delta\boldsymbol{x}$, $\boldsymbol{y} + \Delta\boldsymbol{y}$ and $\boldsymbol{z} + \Delta\boldsymbol{z}$ respectively, obtaining, for $i = 1, \ldots, n$,

$$\boldsymbol{z}_i + \Delta\boldsymbol{z}_i - \left(\|\boldsymbol{z}_i + \Delta\boldsymbol{z}_i\|^2 + \mu^2\right)^{\frac{1}{2}} (\boldsymbol{x}_i + \Delta\boldsymbol{x}_i) = \boldsymbol{0},$$

i.e.

$$\boldsymbol{z}_i + \Delta\boldsymbol{z}_i - \omega_i^\mu \left(1 + 2\frac{\boldsymbol{z}_i^T \Delta\boldsymbol{z}_i}{(\omega_i^\mu)^2} + \frac{\|\Delta\boldsymbol{z}_i\|^2}{(\omega_i^\mu)^2}\right)^{\frac{1}{2}} (\boldsymbol{x}_i + \Delta\boldsymbol{x}_i) = \boldsymbol{0},$$

which gives

$$\boldsymbol{z}_i + \Delta \boldsymbol{z}_i - \omega_i^\mu \left(1 + \frac{\boldsymbol{z}_i^T \Delta \boldsymbol{z}_i}{(\omega_i^\mu)^2} + \frac{\|\Delta \boldsymbol{z}_i\|^2}{2(\omega_i^\mu)^2} - \frac{(\boldsymbol{z}_i^T \Delta \boldsymbol{z}_i)^2}{2(\omega_i^\mu)^4} + \cdots \right) (\boldsymbol{x}_i + \Delta \boldsymbol{x}_i) = \boldsymbol{0}.$$

Moving first-order terms to the left-hand side, constant and second-order terms to the right-hand side, neglecting higher than second-order terms and changing the sign of the equation we obtain

$$(\boldsymbol{E}_\mu \, \Delta \boldsymbol{x} - \boldsymbol{F}_\mu \, \Delta \boldsymbol{z})_i = (\boldsymbol{r}_c)_i - \frac{\boldsymbol{z}_i^T \Delta \boldsymbol{z}_i}{\omega_i^\mu} \Delta \boldsymbol{x}_i - \frac{\|\Delta \boldsymbol{z}_i\|^2}{2\omega_i^\mu} \boldsymbol{x}_i + \frac{(\boldsymbol{z}_i^T \Delta \boldsymbol{z}_i)^2}{2(\omega_i^\mu)^3} \boldsymbol{x}_i \quad (24)$$

for $i = 1, \ldots, n$. The idea, then, is to compute the predictor steps $\Delta \boldsymbol{x}$, $\Delta \boldsymbol{y}$ and $\Delta \boldsymbol{z}$ from equations (15)–(17), and then use these to define the second-order terms which are included in the right-hand side of the linear system solved to obtain the corrector step, using the factorization of $\boldsymbol{A}\boldsymbol{H}_\mu \boldsymbol{A}^T$ a second time.

As noted above, a key component of Mehrotra's method is to exploit the result of the predictor step to define a heuristic value for the centering parameter μ to be used in the computation of the corrector step. This is provided by

$$\tilde{\mu} = \frac{(\operatorname{gap}(\boldsymbol{x} + \Delta \boldsymbol{x}, \boldsymbol{z} + \Delta \boldsymbol{z}))^3}{n\,(\operatorname{gap}(\boldsymbol{x}, \boldsymbol{z}))^2}, \quad (25)$$

where

$$\operatorname{gap}(\boldsymbol{x}, \boldsymbol{z}) = \sum_{i=1}^{n} \left(\|\boldsymbol{z}_i\| - \boldsymbol{x}_i^T \boldsymbol{z}_i \right). \quad (26)$$

This is a natural generalization of Mehrotra's formula for LP. The value $\tilde{\mu}$ may be substituted for μ in *all* the terms on the right-hand side of (24), including the second-order terms as well as the constant term, giving

$$\boldsymbol{h}_i^{(1)} = \boldsymbol{z}_i - \omega_i^{\tilde{\mu}} \boldsymbol{x}_i - \frac{\boldsymbol{z}_i^T \Delta \boldsymbol{z}_i}{\omega_i^{\tilde{\mu}}} \Delta \boldsymbol{x}_i - \frac{\|\Delta \boldsymbol{z}_i\|^2}{2\omega_i^{\tilde{\mu}}} \boldsymbol{x}_i + \frac{(\boldsymbol{z}_i^T \Delta \boldsymbol{z}_i)^2}{2(\omega_i^{\tilde{\mu}})^3} \boldsymbol{x}_i, \quad (27)$$

where

$$\omega_i^{\tilde{\mu}} = \left(\|\boldsymbol{z}_i\|^2 + \tilde{\mu}^2 \right)^{\frac{1}{2}}. \quad (28)$$

It is not practical to substitute $\tilde{\mu}$ for μ on the left-hand side of (24), since the factorization of $\boldsymbol{H}_\mu$ has already been computed using the previous value for μ. Consequently, we also add to the right-hand side of (24) further corrector terms of the form

$$\boldsymbol{h}_i^{(2)} = \left(\omega_i^\mu - \omega_i^{\tilde{\mu}} \right) \Delta \boldsymbol{x}_i + \left(\frac{1}{\omega_i^\mu} - \frac{1}{\omega_i^{\tilde{\mu}}} \right) \left(\boldsymbol{z}_i^T \Delta \boldsymbol{z}_i \right) \boldsymbol{x}_i. \quad (29)$$

As noted in the previous section, the nonsymmetry of $\boldsymbol{H}_\mu$ is a major disadvantage for large sparse problems. We therefore consider here the idea of explicitly symmetrizing $\boldsymbol{H}_\mu$, defining it to be $\frac{1}{2}\boldsymbol{E}_\mu^{-1}(\boldsymbol{F}_\mu + \boldsymbol{F}_\mu^T)$ instead of $\boldsymbol{E}_\mu^{-1}\boldsymbol{F}_\mu$. This suggests subtracting a skew correction term $\frac{1}{2}\boldsymbol{E}_\mu^{-1}(\boldsymbol{F}_\mu - \boldsymbol{F}_\mu^T)\Delta\boldsymbol{z}$ from the right hand side, i.e. adding terms

$$\boldsymbol{h}_i^{(3)} = \frac{1}{2\omega_i^\mu}\left((\Delta\boldsymbol{z}_i^T\boldsymbol{x}_i)\boldsymbol{z}_i - (\Delta\boldsymbol{z}_i^T\boldsymbol{z}_i)\boldsymbol{x}_i\right) \tag{30}$$

to (24). Note the use of ω_i^μ, not $\omega_i^{\tilde{\mu}}$, in the denominator. Putting all this together, the corrector step is defined by

$$\boldsymbol{E}_\mu\,\Delta\boldsymbol{x} - \boldsymbol{F}_\mu\,\Delta\boldsymbol{z} = \boldsymbol{r}_c^c, \tag{31}$$

where the ith block of the "corrected centering" residual is

$$(\boldsymbol{r}_c^c)_i = \begin{cases} \boldsymbol{h}_i^{(1)} + \boldsymbol{h}_i^{(2)} & \text{if} \quad \boldsymbol{H}_\mu = \boldsymbol{E}_\mu^{-1}\boldsymbol{F}_\mu \\ \boldsymbol{h}_i^{(1)} + \boldsymbol{h}_i^{(2)} + \boldsymbol{h}_i^{(3)} & \text{if} \quad \boldsymbol{H}_\mu = \frac{1}{2}\boldsymbol{E}_\mu^{-1}\left(\boldsymbol{F}_\mu + \boldsymbol{F}_\mu^T\right) \end{cases} \tag{32}$$

using (27), (29) and (30).

Substituting $\boldsymbol{r}_c^c$ for $\boldsymbol{r}_c$ in (14), we therefore compute the corrector steps from

$$\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T\Delta\boldsymbol{y} = \boldsymbol{A}\left(\boldsymbol{E}_\mu^{-1}\boldsymbol{r}_c^c + \boldsymbol{x}\right) \tag{33}$$

and

$$\Delta\boldsymbol{x} = \boldsymbol{E}_\mu^{-1}\left(\boldsymbol{F}_\mu\Delta\boldsymbol{z} + \boldsymbol{r}_c^c\right) \tag{34}$$

with $\Delta\boldsymbol{z}$ given by (17).

Note that by analogy with standard practice in LP, it might seem appropriate to modify the right-hand side $\boldsymbol{r}_c$ used by the predictor step by substituting 0 for μ in its definition.[1] In practice, whether or not this is done seems to have little effect, but one reason not to make this choice is that then the dual predictor step is no longer guaranteed to be a descent direction for the smoothed objective function in D_μ. There is no guarantee that the dual corrector step is a descent direction for either this function or the corresponding function defined using $\tilde{\mu}$ instead of μ, although it usually is. If it is a descent direction for the latter function, we update the iterates as before, using $\tilde{\mu}$ instead of μ in the objective function in the dual line search. Otherwise, we abandon both primal and dual corrector steps and use the predictor steps instead.

[1] In fact, this was done in the experiments reported in Section 5.

We now summarize the algorithm. We initialize it with $\boldsymbol{x} = \mathbf{0}$, $\boldsymbol{y}$ set to the minimizer of $\|\boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y}\|$ and $\boldsymbol{z} = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y}$. Assume that an initial value of $\mu > 0$ is given, as well as a termination tolerance ϵ.

Predictor-Corrector Algorithm for Minimizing a Sum of Norms

1. Define ω_i^μ, $\boldsymbol{E}_\mu$ and $\boldsymbol{F}_\mu$ by (12)–(13).
2. Define $\boldsymbol{H}_\mu$ by either $\boldsymbol{E}_\mu^{-1}\boldsymbol{F}_\mu$ or $\frac{1}{2}\boldsymbol{E}_\mu^{-1}(\boldsymbol{F}_\mu + \boldsymbol{F}_\mu^T)$ and find either the LU or Cholesky factorization, respectively, of $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$. In the former case, quit if the LU factorization generates a zero pivot. In the latter case, either quit if the Cholesky factorization fails, or modify the factorization, effectively redefining $\boldsymbol{H}_\mu$ by a nearby positive definite approximation.
3. Determine the predictor steps $\Delta\boldsymbol{y}$, $\Delta\boldsymbol{z}$ and $\Delta\boldsymbol{x}$ from (15), (17) and (16).
4. Define $\tilde{\boldsymbol{x}}$ from the primal scaling rule ((20) and (21)) or the primal steplength rule ((22) and (23)), and $\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{z}}$ by the dual line search rule ((18) and (19)). Quit if the dual line search fails to achieve a reduction in the smoothed dual.
5. Define $\tilde{\mu}$ by (25) and $\omega_i^{\tilde{\mu}}$ by (28).
6. Determine the corrector steps $\Delta\boldsymbol{y}$, $\Delta\boldsymbol{z}$ and $\Delta\boldsymbol{x}$ from (33), (17), (34).
7. If $(\boldsymbol{E}_{\tilde{\mu}}^{-1}\boldsymbol{z})^T\Delta\boldsymbol{z} < 0$, redetermine $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{z}}$ using the primal scaling or steplength rule with $\tilde{\mu}$ instead of μ, and the dual line search rule with $\tilde{\mu}$ instead of μ; quit if the dual line search rule fails to achieve a reduction in the smoothed dual objective.
8. Replace $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ and μ by $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$, $\tilde{\boldsymbol{z}}$ and $\tilde{\mu}$ respectively.
9. If $\|\boldsymbol{A}\boldsymbol{x}\| > \text{gap}(\boldsymbol{x}, \boldsymbol{z})$, quit. (This cannot happen in exact arithmetic and indicates that rounding errors will dominate any further computation.)
10. If $\text{gap}(\boldsymbol{x}, \boldsymbol{z}) < \epsilon$ and $\|c - \boldsymbol{A}^T\boldsymbol{y} - \boldsymbol{z}\| + \|A\boldsymbol{x}\| < \epsilon$, quit; otherwise repeat.

There are several ways the algorithm might terminate when rounding errors prevent further progress: breakdown of the factorization, failure in the line search, or growth in the primal infeasibility $\|\boldsymbol{A}\boldsymbol{x}\|$ with respect to the duality gap measure $\text{gap}(\boldsymbol{x}, \boldsymbol{z})$. The occurrence of any of these conditions essentially indicates that the convergence tolerance ϵ is set too small; in any case, when they occur, the current or previous approximation is generally quite accurate.

4 Experiments on Small Problems

We now report some numerical results for this algorithm, comparing the symmetrized and nonsymmetrized versions and other algorithmic options described above. These results were obtained using a Matlab implementation run on a set of small topologically-constrained Steiner tree test problems (for more details, see [DO98]). The sparsity in the data is determined by the tree structure and its topological constraint, but subject to these qualifications, the data are generated randomly. Each table shows a summary of results from many runs with different random data on the same problem class. Sparsity was not exploited. In all cases $d = 2$. The dual line search was performed using the Matlab **fmin** function with its default tolerance. The machine used was a Sparc Ultra with IEEE double precision arithmetic.

The tables show, for various cases, the number of iterations, the final values of $\text{gap}(\boldsymbol{x}, \boldsymbol{z})$ (defined in (26)) and the infeasibility norm sum $\|\boldsymbol{A}\boldsymbol{x}\| + \|\boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y} - \boldsymbol{z}\|$, each as medians over a set of randomly generated problems in a given class. The termination tolerance ϵ was set to 10^{-10}.

In Tables 1 and 2, we consider a class of Steiner tree problems with $n = 50$, $m = 62$, for which strict complementarity holds at the solution, and for which the median number of indices for which $\|\boldsymbol{z}_i\| = 0$ at the optimal solution is 15. We compare the nonsymmetric and symmetrized variants of the algorithm ($\boldsymbol{H}_\mu = \boldsymbol{E}_\mu^{-1}\boldsymbol{F}_\mu$ and $\boldsymbol{H}_\mu = \frac{1}{2}\boldsymbol{E}_\mu^{-1}(\boldsymbol{F}_\mu + \boldsymbol{F}_\mu^T)$ respectively), with two choices for updating the primal variable: the Scaling Rule and the Steplength Rule with $\tau = 0.999$. All variants used the Dual Line Search Rule. For the symmetrized algorithm, we tested both a version which quits if the Cholesky factorization of $\boldsymbol{H}_\mu$ fails, and one that modifies the factorization and continues iterating: the latter is standard practice in LP [Wri97, p.219]. We also tested a variant of the symmetrized version which omits the skew correction $\boldsymbol{h}_i^{(3)}$. Finally, we also tested the effect of omitting the correction $\boldsymbol{h}_i^{(2)}$, but this had essentially no effect in any case.

Table 1 shows the results for the Primal Scaling Rule and Table 2 shows the results using the Primal Steplength Rule. The notations "skew corr" and "mod Chol" refer to the use of the skew correction term and the modified Cholesky factorization respectively.

The results clearly confirm three remarkable properties of primal-dual

Version	iter (median)	gap (median)	infeas (median)
Not symmetrized	8	$1e-12$	$1e-12$
Symmetrized	8	$1e-06$	$6e-13$
Symmetrized, skew corr	7	$1e-08$	$7e-13$
Symmetrized, mod Chol	15	$5e-11$	$1e-12$
Symmetrized, skew corr, mod Chol	9	$2e-11$	$4e-12$

Table 1: Summary of Results for the Scaling Rule

Version	iter (median)	gap (median)	infeas (median)
Not symmetrized	9	$4e-12$	$6e-13$
Symmetrized	11	$8e-08$	$2e-13$
Symmetrized, skew corr	9	$4e-09$	$1e-14$
Symmetrized, mod Chol	15	$4e-11$	$8e-13$
Symmetrized, skew corr, mod Chol	10	$1e-11$	$4e-13$

Table 2: Summary of Results for the Steplength Rule

predictor-corrector algorithms now well known for linear programming:

- Robust convergence to an optimal solution in all cases tested
- Rapid local convergence so a consistently small number of iterations is required despite the demand for high accuracy
- Highly accurate solutions achieved despite the extremely ill-conditioned linear systems being solved towards the end of the solution process

We now comment in more detail on the results in Tables 1 and 2. First, notice the high accuracy achieved by the nonsymmetric version of the algorithm; the symmetrized version without the modified Cholesky factorization cannot reach the same level of accuracy. With modified Cholesky, high accuracy is achievable, but more iterations are required. The inclusion of the skew correction term $\boldsymbol{h}_i^{(3)}$ substantially improves the performance of the symmetrized version of the algorithm whether or not the Cholesky factorization is modified.

For both the nonsymmetric and symmetrized versions the Primal Scaling Rule has a slightly lower iteration count than the Primal Steplength Rule, apparently because this version of the algorithm has a somewhat faster local

Version	iter (median)	gap (median)	infeas (median)
Not symmetrized	7	$2e-13$	$7e-15$
Symmetrized	16	$6e-11$	$7e-15$
Symmetrized, skew corr	10	$2e-11$	$8e-15$
Symmetrized, mod Chol	16	$6e-11$	$7e-15$
Symmetrized, skew corr, mod Chol	10	$2e-11$	$8e-15$

Table 3: Results for the Scaling Rule on the 20 Chung-Graham ladder problems with strictly complementary solutions

Version	iter (median)	gap (median)	infeas (median)
Not symmetrized	9	$2e-12$	$7e-15$
Symmetrized	16	$6e-11$	$8e-15$
Symmetrized, skew corr	10	$7e-12$	$8e-15$
Symmetrized, mod Chol	16	$6e-11$	$8e-15$
Symmetrized, skew corr, mod Chol	10	$7e-12$	$8e-15$

Table 4: Results for the Steplength Rule on the 20 Chung-Graham ladder problems with strictly complementary solutions

convergence rate. However, we note that the Scaling rule has the significant disadvantage that it is not applicable if nonhomogeneous linear constraints are added to the problem.

In Tables 3 through 6 we display results for a different class of topologically constrained Steiner tree examples, based on the Chung-Graham ladder problem (see [DO98]). For these examples, $n = 85$, $m = 84$, and the median number of indices for which $\|\boldsymbol{z}_i\|$ equals 0 at the optimal solution is 10. For most of these problems, no strictly complementary (SC) solution exists. (Recall from Section 2 that a solution is said to be strictly complementary if, for each i, exactly one of the conditions $\|\boldsymbol{z}_i\| = 0$ and $\|\boldsymbol{x}_i\| = 1$ holds.)

Tables 3 and 4 show results for the 20 cases out of 200 generated where an SC solution is found (using the Primal Scaling and Primal Steplength Rules respectively), while Tables 5 and 6 show results for the other 180 cases where no SC solution is found, presumably because such a solution does not exist. The algorithm achieves the same accuracy (by the duality gap and feasibility measures) on the SC and non-SC problems, but the iteration count is markedly higher in the non-SC case, and the rate of convergence of the algorithm was observed to be slower in the non-SC case. For the non-

Version	iter (median)	gap (median)	infeas (median)
Not symmetrized	14	$3e-11$	$8e-15$
Symmetrized	16	$5e-11$	$8e-15$
Symmetrized, skew corr	14	$4e-11$	$8e-15$
Symmetrized, mod Chol	16	$5e-11$	$8e-15$
Symmetrized, skew corr, mod Chol	14	$4e-11$	$8e-15$

Table 5: Results for the Scaling Rule on the 180 Chung-Graham ladder problems with NO strictly complementary solution

Version	iter (median)	gap (median)	infeas (median)
Not symmetrized	16	$3e-11$	$8e-15$
Symmetrized	22	$5e-11$	$8e-15$
Symmetrized, skew corr	17	$3e-11$	$8e-15$
Symmetrized, mod Chol	22	$5e-11$	$8e-15$
Symmetrized, skew corr, mod Chol	17	$3e-11$	$8e-15$

Table 6: Results for the Steplength Rule on the 180 Chung-Graham ladder problems with NO strictly complementary solution

SC problems, the residuals $\|\boldsymbol{z}_i\|$ are not reduced nearly as close to zero for indices i for which SC does not hold. The reason for this is that the duality gap tolerance requires the products $\boldsymbol{z}_i^T(\boldsymbol{z}_i/\|\boldsymbol{z}_i\| - \boldsymbol{x}_i)$ to be small and *both* factors in the product for such an index i converge to zero as the solution is approached.

For these problems, the modified Cholesky factorization is not needed: the results are identical whether or not it is used.

On the basis of the experiments reported in this section, we recommend the symmetrized version of the algorithm with the skew correction term and the modified Cholesky factorization, using the Dual Line Search and either the Primal Scaling or the Primal Steplength Rule. The choice of the symmetrized version is based on the substantial advantage of being able to use the Cholesky factorization instead of the LU factorization.

5 Large Sparse Problems arising in Plastic Collapse Analysis

A variant of the algorithm described above has been used to solve some challenging large sparse problems arising in plastic collapse analysis. We used a symmetrized version of the algorithm, with $\boldsymbol{H}_\mu = \frac{1}{2}\boldsymbol{E}_\mu^{-1}(\boldsymbol{F}_\mu + \boldsymbol{F}_\mu^T)$, so that the Schur complement $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$ can be factored by Cholesky decomposition, modified to ensure positive definiteness as discussed earlier. This is the primary cost of the algorithm. Details of this and other numerical linear algebra issues are available in [AA97], [AY97] and [And96a]. The primal steplength rule was used, with $\tau = 0.99$ in equation (22).

This sparse implementation was developed over several years with large-scale applications in mind. There are two primary differences from the algorithm discussed in Section 3. The first is that a different generalization of Mehrotra's method was used, based on differentiating a form of the centering condition which incorporates the symmetrization of $\boldsymbol{F}_\mu$ directly, and therefore does not require a skew correction term. The second is that individual centering parameters were used instead of one parameter, namely

$$\tilde{\mu}_i = \begin{cases} \tilde{\mu}, & \text{if} \quad 0.25 < \left(1 - \|\tilde{\boldsymbol{x}}_i\|^2\right) \\ \tilde{\mu}\left(1 - \|\tilde{\boldsymbol{x}}_i\|^2\right)^{-\frac{1}{2}}, & \text{if} \quad \mu \le \left(1 - \|\tilde{\boldsymbol{x}}_i\|^2\right) \le 0.25 \\ \sqrt{\tilde{\mu}}, & \text{if} \quad \left(1 - \|\tilde{\boldsymbol{x}}_i\|^2\right) < \mu \end{cases}$$

for $i = 1, \ldots, n$. This modification was found to give significant improvements in performance for the large-scale problems.

The first three classes of test problems are taken from [And96b], where a primal barrier method was used. We are unaware of any other published results for large sparse problems of the form (D). These test problems are finite dimensional discretizations of collapse problems in rigid plasticity. The discretization step and the physical interpretation of the results can be found in [Chr96] and in [ACO98]. The discrete optimization problems are the same as in [And96b]. The m by dn matrix $\boldsymbol{A}$ is a typical finite element matrix which in plastic analysis is not square since the equilibrium equation for the continuum is under-determined. As earlier, $\boldsymbol{H}_\mu$ is block diagonal with block size $d \times d$. In the cases reported here d is either 2 or 3. The runs were made on the same Convex 3240 vector machine (using IEEE-compatible double precision) as in [And96b] so comparisons of accuracy and CPU time are meaningful.

In the tables, n and m specify the problem dimensions while $|\boldsymbol{A}|$, $|\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T|$

and $|\boldsymbol{L}|$ respectively denote the number of nonzeros in $\boldsymbol{A}$, the upper triangle of $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$ and the Cholesky factor $\boldsymbol{L}$ of $\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T$. These numbers are the same as in [And96b], except for small variations in sparsity due to improvements in the implementation. The iteration count is denoted by "iter" and "cpu" is the CPU time in seconds. The heading "$\|\boldsymbol{z}_i\| = 0$" indicates the number of norms in the dual objective that are zero at the optimal solution. More precisely, $\|\boldsymbol{z}_i\|$ is interpreted as being zero if it is less than the tolerance 10^{-10}. The heading "relgap" denotes the relative duality gap

$$\frac{\left|\sum_{i=1}^{n} \|\boldsymbol{z}_i\| - \boldsymbol{c}^T\boldsymbol{x}\right|}{\sum_{i=1}^{n} \|\boldsymbol{z}_i\| + 1}.$$

In addition to being scaled this is a slightly different measure than the complementarity defined in (26): if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ exactly we have

$$\boldsymbol{c}^T\boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{y} + \boldsymbol{z})^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{z}.$$

Hence the difference is dominated by the primal infeasibility $\|\boldsymbol{A}\boldsymbol{x}\|$ indicated in the last column.

N	n	m	$\|\boldsymbol{A}\|$	$\|\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T\|$	$\|\boldsymbol{L}\|$
4	25	15	224	97	105
10	121	99	1760	1010	1785
50	2601	2499	48800	31010	171083
100	10201	9999	197600	127010	929515
300	90601	89999	1792800	1161010	13203975
400	160801	159999	3190400	2068010	31011299

N	iter	cpu	$\|\boldsymbol{z}_i\| = 0$	relgap	$\|\boldsymbol{A}\boldsymbol{x}\|$
4	7	0	0	$2e-10$	$2e-14$
10	8	1	0	$2e-10$	$7e-15$
50	9	65	0	$2e-09$	$4e-13$
100	10	354	0	$1e-09$	$9e-14$
300	11	6709	0	$1e-09$	$2e-13$
400	12	22139	0	$2e-10$	$2e-13$

Table 7: Problem and solution characteristics for ssp$N$

The first set of problems is denoted ssp$N$ (simply supported plate with a point load solved on an $N \times N$ grid). They all have the same structure, but vary in size, depending on the grid in the finite element analysis. In this problem $d = 3$. This set of problems is characterized by having no zero norms in the solution, i.e., they are, in fact, smooth optimization problems. In [And96b] the constraints $\|\boldsymbol{x}_i\| \leq 1$ are satisfied within a tolerance of order 10^{-9}. These constraints are satisfied exactly in the primal-dual method. Except for this improvement in accuracy, the primal-dual method shows no significant difference, for these problems, compared to the primal barrier method in [And96b]. There is a small reduction in the iteration count, but not in the CPU time. This is clearly a consequence of the fact that these problems are smooth.

For this problem, as well as for the other results reported below, there is no significant difference in the final duality gap and primal infeasibility achieved by the two algorithms. They are, in all cases, about 10^{-8} or less.

The second set of problems, denoted by lNa13, arises in the plane strain model in plasticity. Again N indicates the grid size. In these problems $d = 2$. Characteristics and results are given in Table 8.

N	n	m	$\|\boldsymbol{A}\|$	$\|\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T\|$	$\|\boldsymbol{L}\|$
3	49	52	1390	1142	1207
12	625	640	21331	26406	57421
21	1849	1876	64762	84384	278691
30	3721	3760	131683	175086	726046
60	14641	14720	524403	713766	4337857
99	39601	39732	1425134	1957631	14693151
120	58081	58240	2092843	2881926	24202413

N	iter	cpu	$\|\boldsymbol{z}_i\| = 0$	relgap	$\|\boldsymbol{A}\boldsymbol{x}\|$
3	14	1	12	$5e-10$	$2e-12$
12	19	34	179	$2e-09$	$4e-12$
21	24	200	958	$4e-10$	$4e-12$
30	24	461	2315	$7e-09$	$3e-11$
60	28	4710	11265	$4e-09$	$1e-10$
99	30	17937	33503	$7e-09$	$9e-10$
120	34	44144	50548	$8e-09$	$4e-10$

Table 8: Problem and solution characteristics for lNa13

In this problem set, the number of zero norms varies from 25 percent for $N = 3$, to 87 percent for $N = 120$. Compared with [And96b], there is a significant reduction in the number of iterations and in CPU time. If we disregard the smallest case, $N = 3$, the iteration count for the primal-dual method varies from 19 to 34; for the primal barrier method in [And96b] the variation is from 33 to 176. The primal-dual algorithm also obtains significantly more zero norms in the optimal solution. From our physical understanding of the solution we believe this is correct. It is one of several indications that the primal-dual method is more accurate than the primal barrier method.

There is an important physical interpretation of the complementarity condition (4) in the plasticity problems considered in this section: the vectors $\boldsymbol{z}_i$ represent the deformation (strain) tensor at discrete points in the continuum while the $\boldsymbol{x}_i$ represent the stresses. Thus, if there is any deformation at a point, then the stresses at that point are on their bounds (have norm one) and their directions are determined by the complementarity condition. With this interpretation the complementarity condition is the so-called "flow rule" for the material.

In the third set of test problems, only a small part of the material undergoes deformation; therefore a very large number of the norms are expected to be zero in the optimal solution. As shown in Table 9, the number of zero norms varies from 62 to 96 percent of the total number of terms. Compared with [And96b] the iteration count is significantly reduced and increases very slowly with the problem size. The CPU time is reduced by a factor 4 or more, and we are able to solve larger instances of the problem.

N	n	m	$\lvert\boldsymbol{A}\rvert$	$\lvert\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T\rvert$	$\lvert\boldsymbol{L}\rvert$
20	1681	1718	59054	76945	246012
40	6561	6636	234145	315504	1511239
60	14641	14754	525236	715653	4433304
80	25921	26072	932327	1277402	8645485
120	58081	58308	2094509	2885699	24240011

N	iter	cpu	$\lVert\boldsymbol{z}_i\rVert = 0$	relgap	$\lVert\boldsymbol{A}\boldsymbol{x}\rVert$
20	20	134	1224	$9e-09$	$4e-11$
40	32	1416	6156	$6e-09$	$3e-11$
60	32	4937	14181	$7e-09$	$5e-11$
80	32	12133	25359	$8e-09$	$4e-11$
120	29	36516	57127	$7e-09$	$3e-11$

Table 9: Problem and solution characteristics for lNa20

The last class of test problems is taken from [AC98]. These are problems of the form (D) with additional linear equality constraints:

$$\min\left\{\sum_{i=1}^{n}\|\boldsymbol{z}_i\|\,,\ \text{such that}\ \boldsymbol{A}_i^T\boldsymbol{y}+\boldsymbol{z}_i=\boldsymbol{c}_i, i=1,\ldots,n,\quad \text{and}\ \boldsymbol{E}^T\boldsymbol{y}=\boldsymbol{d}\right\}, \tag{35}$$

where $\boldsymbol{E}\in\Re^{m\times l}$ and $\boldsymbol{d}\in\Re^{l}$, i.e. l is the number of linear constraints. In [AC98], it is shown how the ℓ_1 penalty function approach makes it possible to transform the linearly constrained problem to the unconstrained form (D) in Section 1, and the physical interpretation and setup of the test problems are described. This class of problems is denoted cl$N$13.

N	n	m	l	$\|\boldsymbol{A}\|$	$\|\boldsymbol{A}\boldsymbol{H}_\mu\boldsymbol{A}^T\|$	$\|\boldsymbol{L}\|$
3	9	26	10	163	168	198
12	144	320	145	2818	2790	7402
30	900	1880	901	17848	17511	78128
60	3600	7360	3601	71698	70126	413267
99	9801	19866	9802	195523	191000	1859559
120	14400	29120	14401	287398	280656	2047512
201	40401	81338	40402	807013	787580	9919534
399	159201	319466	159202	3182023	3103949	30001436

N	iter	cpu	$\|\boldsymbol{z}_i\|=0$	relgap	$\|\boldsymbol{A}\boldsymbol{x}\|$	constr
3	11	0	1	$1.6e-08$	$4.4e-09$	$1.0e-15$
12	13	4	95	$6.7e-09$	$2.1e-13$	$1.1e-13$
30	16	31	651	$4.0e-09$	$2.3e-13$	$3.4e-12$
60	20	180	2878	$7.4e-09$	$1.3e-13$	$7.8e-11$
99	24	890	8234	$1.2e-08$	$1.0e-13$	$7.0e-13$
120	25	1238	12311	$1.2e-08$	$1.0e-11$	$2.0e-13$
201	24	6179	35803	$3.1e-08$	$1.0e-13$	$5.1e-13$
399*	35	34776	146326	$7.8e-14$	$2.0e-13$	$6.3e-13$

Table 10: Problem and solution characteristics for cl$N$13.

Characteristics and results for these constrained problems are seen in Table 10. In addition to the number l of linear constraints, there is a new column, "constr", indicating the relative infeasibility of these constraints measured by the expression

$$\frac{\left\| \boldsymbol{E}^T \boldsymbol{y} - \boldsymbol{d} \right\|}{\|\boldsymbol{d}\| + 1}.$$

For the primal barrier method in [AC98], the number of iterations varies from 30 (for $N = 3$ and $N = 12$) to 201 (for $N = 300$). For the primal-dual method the variation is from 11 (for $N = 3$) to 24 (for $N = 201$) and 35 (for $N = 399$). For the case $N = 201$ the CPU time is reduced from 36371 seconds in [AC98] to 6179 using the primal-dual method. However, we can do even better: in the cl$N$13 problems there is one column that is relatively dense, resulting in considerable fill-in during the factorization. Using the technique described in [And96a] for handling dense columns these problems can be solved more efficiently, making it possible to solve for larger values of N. The asterisk in the table indicates that the result for $N = 399$ was obtained by this method. Using the same technique, the case $N = 201$ required 4293 CPU seconds, and there were 6367553 nonzero elements in the $\boldsymbol{L}$ factor.

We conclude that for nonsmooth problems the primal-dual method is significantly more efficient than the primal barrier method applied in [And96b, ACO98, AC98]. The number of iterations increases slowly with the size of the problem. Finally, the primal-dual method appears to be less vulnerable to ill-conditioning near the optimal solution.

Acknowledgement

During the early work on this paper the first author visited the SCCM program of the Computer Science Department at Stanford University. He wishes to thank in particular Michael Saunders for valuable discussions during this period. The fourth author thanks John Michel of Marietta College and JPL for his continued interest in this work.

References

[AA95] I. Adler and F. Alizadeh. Primal-dual interior point algorithms for convex quadratically constrained and semidefinite optimization problems. Technical Report 46-95, RUTCOR, Dec 1995.

[AA97] E.D. Andersen and K.D. Andersen. The APOS linear programming solver: an implementation of the homogeneous algorithm.

Technical Report CORE Discussion Paper 9730, CORE, UCL, Belgium, 1997.

[AC98] K.D. Andersen and E. Christiansen. Minimizing a sum of norms subject to linear equality constraints. *Computational Optimization and Applications*, 1998. To appear.

[ACK+98] J. Alpert, T.F. Chan, A.B. Kahng, I.L. Markov, and P. Mulet. Faster minimization of linear wirelength for global placement. *IEEE Trans. Computer Aided Design for Integrated Circuits and Systems*, 17(1), Jan 1998.

[ACO98] K.D. Andersen, E. Christiansen, and M.L. Overton. Computing limit loads by minimizing a sum of norms. *SIAM Journal on Scientific and Statistical Computing*, 19(3):1046–1062, 1998.

[AM93] J.C. Alexander and J.H. Maddocks. Bounds on the friction-dominated motion of a pushed object. *Int. J. Robotics Research*, 12:231–248, 1993.

[And96a] K. D. Andersen. A modified Schur complement method for handling dense columns in interior-point methods for linear programming. *ACM Transactions on Mathematical Software*, 22(3):348–356, 1996.

[And96b] K.D. Andersen. An efficient Newton barrier method for minimizing a sum of Euclidean norms. *SIAM Journal on Optimization*, 6(1):74–95, 1996.

[AY97] E.D. Andersen and Y. Ye. On a homogeneous algorithm for a monotone complementarity problem with nonlinear equality constraints. In M.C. Ferris and J.S. Pang, editors, *Complementarity and Variational Problems. State of the Art*, pages 1–11. SIAM, Philadelphia, 1997.

[BB] D.V. Byrnes and L.E. Bright. Design of high-accuracy multiple flyby trajectories using constrained optimization. AAS Paper 95-307, AAS/AIAA Astrodynamics Specialist Conference, Halifax, Nova Scotia, Canada, August 14-17, 1995.

[CC80] P.H. Calamai and A.R. Conn. A stable algorithm for solving the multifacility location problem involving Euclidean distances. *SIAM Journal on Scientific and Statistical Computing*, 1:512–526, 1980.

[CC87] P.H. Calamai and A.R. Conn. A projected Newton method for l_p norm location problems. *Mathematical Programming*, 38:75–109, 1987.

[CGM96] T.F. Chan, G.H. Golub, and P. Mulet. A nonlinear primal dual method for TV-based image restoration. In M.Berger et al, editor, *Lecture Notes in Control and Information Sciences*, volume 219, pages 241–252, 1996. Proc. of ICAOS'96, 12 th Int'l Conf. on Analysis and Optimization of Systems: Images, Wavelets and PDE's, Paris, June 26-28, 1996. To appear in SIAM J. Sci. Comp.

[Chr96] E. Christiansen. Limit analysis of collapse states. In P.G. Ciarlet and J.L. Lions, editors, *Handbook of Numerical Analysis*, volume 4, pages 193–312. North–Holland, Amsterdam, 1996.

[DO98] D.R. Dreyer and M.L. Overton. Two heuristics for the Euclidean Steiner tree problem. *Journal on Global Optimization*, 13:95–106, 1998.

[Kuh67] H.W. Kuhn. On a pair of dual nonlinear programs. In J. Abadie, editor, *Nonlinear Programming*, pages 38–54, Amsterdam, 1967. North-Holland.

[Kuh91] H.W. Kuhn. Nonlinear programming: A historical note. In J.K. Lenstra, A.H.G. Rinnooy Kan, and A. Schrijver, editors, *History of Mathematical Programming*, pages 82–96, Amsterdam, 1991. North-Holland.

[Meh92] S. Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, 2:575–601, 1992.

[NT98a] Y. Nesterov and M. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, 8:321–364, 1998.

[NT98b] Y. Nesterov and M. Todd. Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research*, 22:1–42, 1998.

[Ove83] M.L. Overton. A quadratically convergent method for minimizing a sum of Euclidean norms. *Mathematical Programming*, 27:34–63, 1983.

[Ove84] M.L. Overton. Numerical solution of a model problem from collapse load analysis. In J.L. Lions and R. Glowinski, editors, *Computing Methods in Applied Sciences and Engineering VI*, pages 421–437, Amsterdam, 1984. North-Holland.

[Par91] H.R. Parks. Numerical approximation of parametric oriented area-minimizing hypersurfaces. *SIAM Journal on Scientific and Statistical Computing*, 1991.

[Roc70] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton University, 1970.

[Str79] G. Strang. A minimax problem in plasticity theory. In M.Z. Nashed, editor, *Functional Analysis Methods in Numerical Analysis*, New York, 1979. Springer-Verlag. Lecture Notes in Mathematics 701.

[Wei37] E. Weiszfeld. Sur le point par lequel la somme des distances de n points donnes est minimum. *Tohoku Mathematics Journal*, 43:355–386, 1937.

[Wri97] S.J. Wright. *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, 1997.

[XY97] G. Xue and Y. Ye. An efficient algorithm for minimizing a sum of Euclidean norms with applications. *SIAM Journal on Optimization*, 7:1017–1036, 1997.