A Primal-Dual Algorithm for Minimizing a Non-Convex Function Subject to Bound and Linear Equality Constraints

A R Conn N I M Gould and P L Toint

November 1996
A primal-dual algorithm for minimizing a non-convex function subject to bound and linear equality constraints

Andrew R. Conn¹, Nicholas I. M. Gould²,³ and Philippe L. Toint⁴,⁵

ABSTRACT
A new primal-dual algorithm is proposed for the minimization of non-convex objective functions subject to simple bounds and linear equality constraints. The method alternates between a classical primal-dual step and a Newton-like step in order to ensure descent on a suitable merit function. Convergence of a well-defined subsequence of iterates is proved from arbitrary starting points. Algorithmic variants are discussed and preliminary numerical results presented.

¹ IBM T.J. Watson Research Center, P.O.Box 218, Yorktown Heights, NY 10598, USA
   Email: arconn@watson.ibm.com

² Department for Computation and Information, Rutherford Appleton Laboratory,
   Chilton, Oxfordshire, OX11 0QX, England, EU
   Email: nimg@letterbox.rl.ac.uk

³ Current reports available by anonymous ftp from joyous-gard.cc.rl.ac.uk
   (internet 130.246.9.91) in the directory “pub/reports”.

⁴ Department of Mathematics, Facultés Universitaires ND de la Paix,
   61, rue de Bruxelles, B-5000 Namur, Belgium, EU
   Email: pht@math.fundp.ac.be

⁵ Current reports available by anonymous ftp from thales.math.fundp.ac.be
   (internet 138.48.20.102) in the directory “pub/reports”.

Department for Computation and Information
Atlas Centre
Rutherford Appleton Laboratory
Oxon OX11 0QX
November 19, 1996.
1 Introduction: the problem and the algorithm

1.1 The problem

In this paper, we consider algorithms for solving general (i.e., non-convex), linearly constrained, differentiable optimization problems. We shall distinguish between simple bounds and general linear constraints, and find it convenient to reformulate inequalities as equalities via slack variables. We thus consider the problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{s.t.} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

where \( f(\cdot) \) is a real valued function on \( \mathbb{R}^n \), \( x \) is a vector in \( \mathbb{R}^n \), \( A \) is an \( m \times n \) matrix and \( b \) is a vector of \( \mathbb{R}^m \).

In part, we are motivated to consider the above problem because of our experiences with the general large-scale nonlinear programming package LANCELOT (Conn, Gould and Toint, 1992). In this package, simple bounds are treated explicitly and all other constraints are converted to equations and incorporated into an augmented Lagrangian merit function. While this proves to be a robust approach (Conn, Gould and Toint, 1996), it has a number of obvious drawbacks. One of these is that augmentation may not be the ideal way to treat linear constraints, and a more attractive approach is to handle all linear constraints explicitly (Conn, Gould, Sarketaer and Toint, 1996). We note that there has been a relatively long history of methods that use linearly constrained subproblems at their heart. References include the methods of Rosen and Kreuser (1972), Robinson (1972), and Murtagh and Saunders (1978), the latter being the basis of the well-known large-scale nonlinear programming package MINOS.

Another drawback with the LANCELOT approach is the use of the simple bounds that are active at the generalized Cauchy point to predict those which will be active at the solution (see the trust region based kernel algorithm SMIN, Conn, Gould and Toint, 1988). Unfortunately this approach does not appear to be very effective when the problem is either degenerate or close to degenerate. On the other hand interior point methods, particularly primal-dual approaches, have enjoyed much success in linear programming and it is generally accepted that any state-of-the-art library for linear programming should include both interior point and simplex methods (for example OSL from IBM, 1990, and CPLEX, 1995). It is usually acknowledged that interior point methods are less sensitive to degeneracy than active set methods, see for example Shanno (1994). Thus we were motivated to consider an interior point method in which linear constraints \( Ax = b \) are handled explicitly and simple bounds are handled via a logarithmic barrier term. For the record, we still expect to handle general nonlinear constraints using the augmented Lagrangian. However, we do want to retain the flexibility of not necessarily satisfy the linear constraints during the earlier iterations.

In addition, since the linear programming problem is a convex linear problem, it is the case that the first order conditions are sufficient to characterize a solution and thus it is possible to dispense with a merit function entirely. In the non-convex case, the merit function is an essential ingredient of any successful algorithm and the choice of merit function was a considerable concern in the present paper.

However noble one may believe these goals, there are some significant difficulties in an in-
terior point approach. Besides those already mentioned there is an additional discussion in the conclusions of this paper. Although we are not successful in addressing all these issues, and indeed some of the most important practical issues will depend upon much more extensive testing, what we do hope we have achieved in the present paper is a consistent method with a single merit function and a guaranteed descent direction that either is the primal dual direction or a very closely related Newton direction. In addition, linear equalities are treated explicitly without requiring primal feasibility initially.

Considering the vast literature on primal-dual methods for convex problems, there has been remarkably little work on extending these methods to the non-convex case. This may be because dual variables are not globally meaningful for non-convex problems, but one is tempted to believe that in the neighbourhood of a minimizer some sort of local convexity may be amenable to a primal-dual approach. Indeed, Simantiraki and Shanno (1995) analyse such a local method. Globally, of course, one would expect to require a merit function to force convergence, and Forsgren and Gill (1996) provide just such a function for primal-dual methods. A complete analysis of an interior-point algorithm for non-convex linearly constrained optimization is provided by Bonnans and Pola (1993), but this algorithm appears to require both a strictly interior starting point and a convex model of the objective.

Although the emphasis here is on theoretical issues, we do include preliminary results on a non-trivial set of general quadratic programming problems from the the CUTE test set (see, Bongartz, Conn, Gould and Toint, 1995) which we compare with a state-of-the-art active set method designed for solving quadratic programs. Before going into further details of the proposed algorithm we include some additional notation and our assumptions.

If we denote the Euclidean inner product by $(\cdot, \cdot)$ and let $e$ be the vector of all ones, we assume that

AS1. $f(\cdot)$ is a twice continuously differentiable,

AS2. the function $f(x) - \mu(\log(x), e)$ is bounded below on the positive orthant for every $\mu > 0$,

AS3. the gradient and Hessian of $f(x)$ are uniformly bounded in norm over the positive orthant,

AS4. $A$ has full rank, and

AS5. there exists a point $x_\circ$ strictly interior to the positive orthant such that $Ax_\circ = b$.

### 1.2 The primal-dual search direction

The first order criticality conditions for problem (1.1) may be written as

\[
\begin{align*}
g(x) + A^T y - z &= 0, \\
Az &= b, \\
XZe &= 0, \\
(x, z) &\geq 0,
\end{align*}
\]

(1.2)

where $z$ is a vector of $\mathbb{R}^n$, $g(z) \equiv \nabla_x f(x)$ and

\[
X = \text{diag}(x_1, \ldots, x_n) \text{ and } Z = \text{diag}(z_1, \ldots, z_n).
\]

(1.3)
In order to build our algorithm, we consider a perturbed version of this system of equations given by

\[
\begin{align*}
g(x) + ATy - z &= 0 \\
Ax &= b \\
XZc &= \mu c, \\
(x, z) &\geq 0,
\end{align*}
\]

where

\[
\mu = \sigma \frac{(x, z)}{n},
\]

for some given \( \sigma \in (0, 1) \). Our algorithm moves from the current estimate \((x_k, z_k) > 0\) of the solution of (1.1) to a new estimate \((x_{k+1}, z_{k+1}) > 0\) given by

\[
x_{k+1} = x_k + \alpha_k^{(x)} \Delta x_k \quad \text{and} \quad z_{k+1} = z_k + \alpha_k^{(z)} \Delta z_k,
\]

for some scalar stepsizes \(\alpha_k^{(x)}, \alpha_k^{(z)} \in (0, 1)\), where \(\Delta x_k\) and \(\Delta z_k\) may, for instance, be chosen as \(\Delta x_k^{PD}\) and \(\Delta z_k^{PD}\) which solve the system

\[
\begin{align*}
H_k &\Delta x_k^{PD} + ATy_{k+1}^{PD} - \Delta z_k^{PD} &= -g_k + z_k, \\
A &\Delta x_k^{PD} = b - Ax_k \\
Z_k &\Delta x_k^{PD} + X_k \Delta z_k^{PD} = \mu_k c - X_k Z_k c,
\end{align*}
\]

where \(H_k \equiv H(x_k) \equiv \nabla_{xx} f(x_k)\) and where \(g_k \equiv g(x_k)\). This system is a linearization, at \((x_k, z_k)\), of (1.4), in which \(y_{k+1}\) is considered as an auxiliary variable. Eliminating \(\Delta z_k^{PD}\), and defining

\[
r_k = Ax_k - b,
\]

we obtain that

\[
\begin{pmatrix}
H_k + X_k^{-1} Z_k & AT \\
A & 0
\end{pmatrix}
\begin{pmatrix}
\Delta x_k^{PD} \\
y_{k+1}^{PD}
\end{pmatrix}
= - \begin{pmatrix}
g_k - \mu_k X_k^{-1} c \\
r_k
\end{pmatrix}
\]

and

\[
\Delta z_k^{PD} = -z_k - X_k^{-1} Z_k \Delta x_k^{PD} + \mu_k X_k^{-1} c.
\]

Note that (1.9) fully defines \(\Delta x_k^{PD}\), and \(y_{k+1}^{PD}\) provided the matrix \(H_k + X_k^{-1} Z_k \equiv G_k\) is nonsingular in the nullspace of \(A\). This is obviously the case if \(f(x)\) is strictly convex, but may not be true in general. We discuss below how \(G_k\) might be modified or how \(\Delta x_k^{PD}\) may be defined in more general situations. Observe also that, if this quantity is well defined, \(\Delta z_k^{PD}\) is in turn well defined by (1.10). The strict positivity of \(x_{k+1}\) and \(z_{k+1}\) is ensured by suitably restricting the stepsizes \(\alpha_k^{(x)}\) and \(\alpha_k^{(z)}\), as is detailed below. Thus, if at the solution \(x_*\) or \(z_*\) have zero components, these can only be attained in the limit.

Observe that we may now introduce an artificial variable \(\xi\) in the system (1.9), which is defined by

\[
Ax - b = \xi r_0,
\]

which is possible for a scalar variable because of the second equation of (1.7). If \(r_0 \neq 0\), this artificial variable is initially one; at each iteration, we augment the primal-dual step with the correction

\[
\Delta \xi_k^{PD} = -\xi_k
\]

to \(\xi_k\). Thus if a unit step is ever taken, the linear equality constraints will be satisfied exactly from then on. We will use the notation \(v = (x, \xi)\) to denote points in the \((x, \xi)\)-space.
1.3 An alternative search direction

When $\xi > 0$, we may now consider the alternative problem of minimizing the shifted penalty function

$$f(x) + \frac{1}{2} \rho(\xi + 1)^2$$

subject to the constraints (1.11) and

$$x \geq 0.$$  

In this formulation, the shifted penalty terms drives the variable $\xi$ below zero for sufficiently large $\rho$. We then intend to stop the minimization prematurely as soon as $\xi$ attains the value zero.

Writing the first order optimality conditions for this modified problem, we obtain that

$$
geq g(x) + A^T y - z = 0, \\
-\langle r_0, y \rangle + \rho(\xi + 1) = 0, \\
Ax - \xi r_0 = b, \\
XE = 0, \\
(x, z) \geq 0. \tag{1.15}$$

We perturb the system in the same manner as above and write the corresponding Newton’s iteration, which yields that

$$H_k \Delta x_k^N + A^T y_{k+1}^N - \Delta x_k^N = g_k + z_k, \\
A \Delta x_k^N - \Delta \xi_k^N r_0 = 0, \\
-\langle r_0, y_{k+1}^N \rangle + \rho_k \Delta \xi_k^N = -\rho_k(\xi_k + 1), \\
Z_k \Delta x_k^N + X_k \Delta z_k = \mu_k e - X_k Z_k e. \tag{1.16}$$

As before, we may eliminate $\Delta z_k^N$, and obtain that

$$
\begin{pmatrix}
H_k + X_k^{-1} Z_k & A^T & 0 \\
A & 0 & -r_0 \\
0 & -r_0^T & \rho_k \\
\end{pmatrix}
\begin{pmatrix}
\Delta x_k^N \\
y_{k+1}^N \\
\Delta \xi_k \\
\end{pmatrix}
= -
\begin{pmatrix}
g_k - \mu_k X_k^{-1} e \\
0 \\
\rho_k(\xi_k + 1) \\
\end{pmatrix}. \tag{1.17}
$$

and

$$\Delta z_k^N = -z_k - X_k^{-1} Z_k \Delta x_k^N + \mu_k X_k^{-1} e. \tag{1.18}$$

1.4 The merit function

We now introduce, for given $\mu, \rho > 0$, the logarithmic penalty function defined by

$$\phi(v, \mu, \rho) = f(x) + \frac{1}{2} \rho(\xi + 1)^2 - \mu(\log(x), e). \tag{1.19}$$

Examining now the derivative of this function, we find that

$$\nabla_x \phi(v, \mu, \rho) = g(x) - \mu X^{-1} e \quad \text{and} \quad \nabla_\xi \phi(v, \mu, \rho) = \rho(\xi + 1). \tag{1.20}$$

We first consider the slope of this function at a given iterate $v_k^T$ along the step

$$\Delta v_k^N = ((\Delta x_k^N)^T, \Delta \xi_k^N). \tag{1.21}$$
defined by (1.16) (or, equivalently, (1.17)), and we obtain that

\[
\langle \nabla_y \phi_k, \Delta u_k^N \rangle = - \left( (\Delta x_k^N)^T (y_{k+1}^N)^T \Delta \xi_k^N \right) \begin{pmatrix} G_k & A^T & 0 \\ A^T & 0 & -r_0 \\ 0 & -r_0^T & \rho_k \end{pmatrix} \begin{pmatrix} \Delta x_k^N \\ y_{k+1}^N \\ \Delta \xi_k^N \end{pmatrix},
\]

(1.22)

where we have defined \( \phi_k = \phi(u_k, \mu_k, \rho_k) \). Using the second equation of (1.17), (1.22) gives that

\[
\langle \nabla_y \phi_k, \Delta u_k^N \rangle = -(\Delta x_k^N, G_k \Delta x_k^N) - \rho_k (\Delta \xi_k^N)^2.
\]

(1.23)

On the other hand, the direction

\[
(\Delta u_k^{PD})^T = ((\Delta x_k^{PD})^T, \Delta \xi_k^{PD}),
\]

(1.24)

defined by (1.7) (or, equivalently, (1.9)) and (1.12), yields the slope

\[
\langle \nabla_y \phi_k, \Delta u_k^{PD} \rangle = \langle \Delta x_k^{PD}, g_k - \mu_k X_k^{-1} e \rangle - \rho_k \xi_k (\xi_k + 1)
\]

\[
= -(\Delta x_k^{PD}, G_k \Delta x_k^{PD}) - (\Delta x_k^{PD}, A^T y_{k+1}^{PD}) - \rho_k \xi_k (\xi_k + 1)
\]

\[
= -(\Delta x_k^{PD}, G_k \Delta x_k^{PD}) - \rho_k (\Delta \xi_k^{PD})^2 - \xi_k (\rho_k - (r_0, y_{k+1}^{PD}))
\]

(1.25)

where we have used (1.7), the definition of \( \xi \) and (1.12).

We now examine under which conditions the slopes given by (1.23) and (1.25) are negative. To this aim, we introduce the following definition: the matrix \( G \) is said to be second-order sufficient with respect to \( A \) if and only if the augmented matrix

\[
K = \begin{pmatrix} G & A^T \\ A & 0 \end{pmatrix}
\]

(1.26)

is nonsingular and has precisely \( m \) negative eigenvalues. This is equivalent to requiring that \( (y, Gy) > 0 \) for all nonzero \( y \) satisfying \( Ay = 0 \), or to the reduced matrix \( N^T G N \) being positive definite, where the columns of \( N \) span the nullspace of \( A \) (see, for instance, Gould, 1985). The matrix is second-order necessary if we drop the requirement that \( K \) be nonsingular; this is then equivalent to requiring that \( (y, Gy) \geq 0 \) for all \( y \) satisfying \( Ay = 0 \) or to the reduced matrix \( N^T G N \) being positive semidefinite. Thus, returning to (1.23), we see that we have descent with the Newton direction if we insist that \( G_k \) be second-order sufficient.

If \( \xi_k = 0 \) then the identity

\[
\xi_k r_0 = r_k
\]

(1.27)

and (1.9) gives that \( A \Delta x_k = 0 \). Thus, if the matrix \( G_k \) is second-order sufficient with respect to \( A \), we may deduce that

\[
\langle \nabla_y \phi_k, \Delta u_k^{PD} \rangle = -(\Delta x_k^{PD}, G_k \Delta x_k^{PD}) < 0
\]

(1.28)

If we now consider the case where \( G_k \) is second-order sufficient with respect to \( A \) but \( \xi_k \neq 0 \), it turns out that we can still show that the slope (1.25) is negative provided we choose \( \rho_k \) large enough. This result from the two following lemmas.

**Lemma 1** Assume that the matrix \( G \) is second-order sufficient with respect to \( A \) and that the columns of \( N \) are orthogonal. Then the smallest eigenvalue of \( N^T G N \) is at least equal to the smallest positive eigenvalue of \( K \), where \( K \) is defined by (1.26).
Proof. If $G$ is second-order sufficient with respect to $A$ and $N$ is orthogonal, we have that the minimum eigenvalue of $N^TGN$, which we denote by $\epsilon > 0$, is the solution of the minimization problem
\[
\min_s \{(s, Gs) \mid As = 0 \text{ and } \|s\| = 1\}. \tag{1.29}
\]
(Here and below, the symbol $\| \cdot \|$ denotes the Euclidean norm.) The minimizer of this problem satisfies the first-order optimality conditions
\[
Gs + A^T u = \epsilon s \\
As = 0 \tag{1.30}
\]
and $\epsilon = (s, Gs)$. Adding $eu$ on both sides of the second equation, we see that (1.30) yields that
\[
\begin{pmatrix} G & A^T \\ A & \epsilon I \end{pmatrix} \begin{pmatrix} s \\ u \end{pmatrix} = \epsilon \begin{pmatrix} s \\ u \end{pmatrix}, \tag{1.31}
\]
and thus $\epsilon$ is an eigenvalue of the matrix
\[
\begin{pmatrix} G & A^T \\ A & \epsilon I \end{pmatrix}. \tag{1.32}
\]
Now, we can view the matrix $K$ defined in (1.26) as a symmetric perturbation of (1.32), and deduce from Wilkinson (1965, Section 44, p. 101), that $K$ has an eigenvalue in the range $[0, \epsilon]$. Since $K$ is nonsingular, this eigenvalue must be in the interval $(0, \epsilon]$, which proves the result. \[\square\]

Lemma 2 \[1\] Assume that the matrix $G$ is second-order sufficient with respect to $A$, and that the smallest strictly positive eigenvalue of $K$ is $\lambda > 0$. Then, if one chooses an arbitrary $m$-dimensional vector $r$ and if
\[
\rho \geq \lambda + \frac{2}{\lambda} \|r\|^2, \tag{1.33}
\]
the matrix
\[
\tilde{G} = \begin{pmatrix} G & 0 \\ 0 & \rho \end{pmatrix} \tag{1.34}
\]

\[1\] It is interesting to note that Lemma 2 does not hold if second-order sufficiency is replaced by second-order necessity. For, suppose that
\[
G = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad r = -1.
\]
Then the columns of the matrix
\[
N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]
form a basis of the nullspace of $(A r)$, and the resulting "reduced matrix" is
\[
N^T \begin{pmatrix} G & 0 \\ 0 & \rho \end{pmatrix} N = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 + \rho \end{pmatrix}.
\]
Unfortunately, this latter matrix has a negative eigenvalue for all $\rho$. 

6
is second-order sufficient with respect to \((A \tau)\) and
\[
\langle v, \tilde{G}v \rangle \geq \frac{1}{2} \lambda \|v\|^2
\]  
(1.35)
for \(v = (x, \xi)\) in the nullspace of \((A \tau)\).

**Proof.** Consider the matrix
\[
\tilde{K}_1 = \begin{pmatrix}
G & 0 & A^T \\
0 & \rho - \lambda & r^T \\
A & r & 0
\end{pmatrix}
\]  
(1.36)
Pivoting on the 2-2 block and using Sylvester’s law of inertia, we obtain that
\[
\text{In}(\tilde{K}_1) = (1, 0, 0) + \text{In}(K(\rho)), \text{ where } K(\rho) \overset{\text{def}}{=} \begin{pmatrix}
G & A^T \\
A & -rr^T/(\rho - \lambda)
\end{pmatrix}.
\]  
(1.37)
As, by assumption, \(K\) is nonsingular and has exactly \(m\) negative eigenvalues, Wilkinson (1965, Section 40, p. 97) implies that the smallest positive eigenvalue of \(K(\rho)\) is at least \(\frac{1}{2} \lambda\) provided that
\[
\frac{\|rr^T\|}{\rho - \lambda} = \frac{\|r\|^2}{\rho - \lambda} \leq \frac{1}{2} \lambda,
\]  
(1.38)
i.e., provided \(\rho\) satisfies (1.33). The continuity of the eigenvalues of \(K(\rho)\) then implies that both \(K(\rho)\), and, in view of (1.37), \(\tilde{K}_1\), also have precisely \(m\) negative eigenvalues for all \(\rho\) satisfying (1.33). Thus \(\tilde{N}^T \tilde{G}_1 \tilde{N}\) is positive definite118 with \(\tilde{G}_1 = \text{diag}(G, \rho - \lambda)\) and where the columns of \(\tilde{N}\) span the nullspace of \((A \tau)\). As a consequence,
\[
\tilde{N}^T \tilde{G}_1 \tilde{N} = \tilde{N}^T (\tilde{G}_1 + \text{diag}(0, \lambda)) \tilde{N}
\]  
(1.39)
is also positive definite, which proves the first part of the lemma.

To prove the second part, observe that
\[
\tilde{K} = \begin{pmatrix}
G & 0 & A^T \\
0 & \lambda & 0 \\
A & 0 & -rr^T/(\rho - \lambda)
\end{pmatrix} + \frac{1}{\rho - \lambda} \begin{pmatrix}
0 & \rho - \lambda & r^T \\
0 & \rho - \lambda & r^T \\
A & 0 & -rr^T/(\rho - \lambda)
\end{pmatrix} \overset{\text{def}}{=} K_1 + K_2.
\]  
(1.40)
But the eigenvalues of \(K_1\) are \(\lambda\) and those of \(K(\rho)\): the smallest positive eigenvalue of \(K_1\) is thus at least \(\frac{1}{2} \lambda\) so long as (1.33) holds. Moreover, \(K_2\) is a positive rank-one term, which implies that the eigenvalues of \(\tilde{K}\) are not smaller than those of \(K_1\). Recalling that \(\tilde{K}\) has exactly \(m\) negative eigenvalues if (1.33) holds, we see that its smallest positive eigenvalue is at least \(\frac{1}{2} \lambda\). Applying now Lemma 1 gives (1.35). □

Returning to the sign of the slopes of (1.23) and (1.25) in the case where \(G_k\) is second-order sufficient with respect to \(A\) and \(\xi_k \neq 0\), we see immediately, from (1.23), Lemma 2 and the requirement \(A \Delta x_k^N - \Delta \xi_k^N r_0 = 0\), that
\[
\langle \nabla_v \phi_k, \Delta \xi_k^N \rangle \leq -\frac{1}{2} \lambda (\| \Delta x_k^N \|^2 + (\Delta \xi_k^N)^2).
\]  
(1.41)
so long as
\[
\rho_k \geq \lambda + \frac{2}{\lambda} \|r_0\|^2,
\]  
(1.42)
where $\lambda$ is the smallest eigenvalue of $N^T G_k N$. We also see that the second equation of (1.9) can be rewritten as

$$A\Delta x_k^{PD} - \Delta \xi_k r_0 = 0,$$

(1.43)

and thus may deduce from (1.25), Lemma 2, (1.12), and (1.43) that

$$\langle \nabla v \phi_k, \Delta v_k^{PD} \rangle \leq -\frac{1}{2} (\|\Delta x_k^{PD}\|^2 + (\Delta \xi_k^{PD})^2).$$

(1.44)

whenever

$$\rho_k \geq \max \left\{ \lambda + \frac{2}{\lambda} \|r_0\|^2, (r_0, y_{k+1}^{PD}) \right\}.$$  

(1.45)

Observe that condition (1.45) depends on the size of $\langle r_0, y_{k+1}^{PD} \rangle$. The penalty parameter $\rho_k$ may thus become too large because of this latter quantity, in which case we might prefer to use the alternative formulation using the shifted quadratic penalty term for which descent is always obtained (see (1.41)) if $G_k$ is second order sufficient with respect to $A$ (see (1.23)). Our algorithm takes advantage of this observation.

1.5 Modifications

If $G_k$ is not second-order sufficient with respect to $A$, we may add a positive semidefinite modification $\Delta G_k$ to $G_k$, so that $G_k + \Delta G_k$ is uniformly second-order sufficient with respect to $A$, meaning that the minimum eigenvalue of $N^T (G_k + \Delta G_k) N$ is larger than some $\lambda > 0$ independent of $k$. In turn yields well defined $\Delta x_k^{PD}$ and $y_{k+1}^{PD}$, and ensures (1.28). The smallest such modification may need to be as large as $\|N^T G_k N\| + \lambda$, but here we merely require that

$$\|\Delta G_k\| \leq \kappa_2 \|G_k\| + \lambda$$

(1.46)

for some $\kappa_2 \geq 1$. The modification $\Delta G_k$ to make $N^T G_k N$ positive definite may be much smaller than that required to make $G_k$ itself positive definite.

The technique of ensuring the second-order sufficiency of $G_k$ with respect to $A$ is not the only one which can be considered to make the slope (1.25) negative. One could also modify $\Delta x_k^{PD}$ to include a sufficient contribution of a direction of negative curvature, provided the second equation of (1.9) remains satisfied.

The fact that the directional derivative (1.28) is negative ensures that the (possibly modified) primal-dual step $\Delta v_k^{PD}$ is a descent direction for $\phi_k$, when $v_k$ is not a minimizer. We may thus consider using this function as a “merit function” associated with this step, that is with the linearization of conditions (1.4).

The viability of such approaches are discussed further, with additional references in Forsgren and Murray (1993), Gould (1995) and Higham and Cheng (1996).

1.6 The step

We now turn to the question of determining the stepsizes in (1.6). A first and crucial constraint on the stepsizes is induced by our decision to maintain both $z_{k+1}$ and $z_{k+1}$ strictly positive. We thus have to specify some bounds on $\alpha_k (x)$ and $\alpha_k (z)$ that will guarantee that the iterates remains “sufficiently” inside the positive orthant of the $(x, z)$-space. When both stepsizes are chosen equal
(i.e. $\alpha_k^{(x)} = \alpha_k^{(z)})$, a set of suitable conditions (see Simantiraki and Shanno (1995) or Zhang and Zhang (1994)) on the (unique) stepsize is given by the inequalities

$$X_{k+1}Z_{k+1}e \geq \gamma \frac{\langle x_{k+1}, z_{k+1} \rangle}{n} e \tag{1.47}$$

and

$$\langle x_{k+1}, z_{k+1} \rangle \geq \gamma \| g_{k+1} + A^T y_{k+1} - z_{k+1} \|,$$ \hspace{1cm} \tag{1.48}

where $\gamma \in (0, 1)$. We observe that conditions (1.47) and (1.48) clearly ensure that $x_{k+1}$ and $z_{k+1}$ both have all components strictly positive so long as the conditions

$$z_{k+1} = g_{k+1} + A^T y_{k+1} \quad \text{and} \quad \langle x_{k+1}, z_{k+1} \rangle = 0 \tag{1.49}$$

are violated. On the other hand, condition (1.47) and (1.48) appear to be somewhat restrictive in practice because (1.47) often restricts the step in $x$ more than necessary. We might thus prefer to keep independent stepsizes in $x$ and $z$ and require, instead of (1.47), that

$$X_{k+1}e \geq \omega(\mu_k)X_ke$$ \hspace{1cm} \tag{1.50}

where $\omega(\mu_k) \in (0, 1)$ is a small parameter possibly dependent on the value of $\mu_k$. Note that the largest stepsize ensuring (1.50) is given by

$$\tilde{\alpha}_k^{(x)} = \min \left[ 1, \frac{\omega(\mu_k)}{1 - \omega(\mu_k)} \min_{\Delta z_i < 0} \frac{-\Delta z_i}{\Delta z_i} \right], \hspace{1cm} \tag{1.51}$$

where $\lceil u \rceil_i$ denotes the $i$-th component of the vector $u$. However, if this maximum stepsize is adequate for the primal-dual step $\Delta w_k^{PD}$ in that (1.12) ensures that

$$\xi_k + \tilde{\alpha}_k^{(x)} \Delta z_k^{PD} \geq 0,$$ \hspace{1cm} \tag{1.52}

this may not be the case for the Newton step $\Delta v_k^N$ because $\Delta z_k^N$ is now defined from the solution of (1.16). Indeed, for $\rho_k$ large enough, we would expect $\xi_k$ to tend to $-1$. We thus have to limit the stepsize to maintain $\xi_{k+1}$ non-negative: the largest stepsize in $\xi$ is now given by

$$\tilde{\alpha}_k^{(x)} = \begin{cases} \min[1, -\xi_k/\Delta z_k^N] & \text{if } \Delta z_k^N < 0, \\ 1 & \text{otherwise}. \end{cases} \tag{1.53}$$

(Note that a zero value of $\xi_k$ is desirable, as it implies primal feasibility of the iterates.) Combining these bounds, we obtain that the maximum stepsize in the $v = (x, \xi)$ space is given by

$$\tilde{\alpha}_k^{(v)} = \begin{cases} \tilde{\alpha}_k^{(x)} & \text{if } \Delta v_k = \Delta v_k^{PD}, \\ \min[\tilde{\alpha}_k^{(x)}, \tilde{\alpha}_k^{(\xi)}] & \text{if } \Delta v_k = \Delta v_k^N. \end{cases} \tag{1.54}$$

We may then calculate the actual stepsize

$$\alpha_k^{(v)} = \beta^j \tilde{\alpha}_k^{(v)},$$ \hspace{1cm} \tag{1.55}$$

by a classical Armijo linesearch procedure, that is by determining the smallest nonnegative integer $j_k$ such that

$$\phi(v_k + \beta^j \tilde{\alpha}_k^{(v)} \Delta v_k, \mu_k, \rho_k) \leq \phi_k + \eta \beta^j \tilde{\alpha}_k^{(v)} \langle \nabla \phi_k, \Delta v_k \rangle$$ \hspace{1cm} \tag{1.56}$$

for some $\beta \in (0, 1)$.
1.7 The algorithm

We are now in position to formally state our algorithm.

**Algorithm**

**Step 0:** Set $k = 0$. The starting iterate $(x_0, 1, z_0)$ is given, such that $x_0, z_0 > 0$, as well as the initial barrier parameter $\mu_0 > 0$ and the constants $0 < \beta, \lambda, \eta, \nu_1, \sigma, \omega < 1$, $\theta^\text{DP}, \theta^\text{FF} > 0$, $\delta, \nu_2 > 1$, $\theta^C \in (1, 1/\sigma)$, and $\rho_0 \geq \rho_{\min} \overset{\text{def}}{=} \lambda + 2 \|r_0\|/\lambda$. Define $\xi_0 = 1$ and set $\omega(\mu_0) \in (0, \bar{\omega})$.

**Step 1:** Compute the primal-dual step $\Delta v^\text{PD}_k$ and $y^\text{PD}_{k+1}$ from (1.9) and (1.12), modifying $\bar{G}_k$ if necessary to ensure that it is uniformly second-order sufficient with respect to $A$ (with constant $\lambda$).

**Step 2:** If either $\xi_k = 0$ or (1.44) holds, define $\Delta v_k = \Delta v^\text{PD}_k$ and $y_{k+1} = y^\text{PD}_{k+1}$. Otherwise, compute the Newton step $\Delta v^N_k$ and $y^N_{k+1}$ from (1.17) and set $\Delta v_k = \Delta v^N_k$ and $y_{k+1} = y^N_{k+1}$.

**Step 3:** Compute $\alpha^{(v)}_k$ from (1.55) and (1.56). Then set

$$x_{k+1} = x_k + \alpha^{(v)}_k \Delta x_k \quad \text{and} \quad \xi_{k+1} = \xi_k + \alpha^{(v)}_k \Delta \xi_k. \quad (1.57)$$

**Step 4:** Define

$$\Delta z_k = -z_k - X^{-1}_k Z_k \Delta x_k + \mu_k X_k^{-1} e. \quad (1.58)$$

If $z_k + \Delta z_k$ lies (componentwise) in the interval

$$\left[\nu_1 \min(e, z_k, \mu_k X_k^{-1} e), \max(\nu_2 e, z_k, \nu_2 \mu_k X_k^{-1} e, \nu_2 \mu_k X_k^{-1} e)\right], \quad (1.59)$$

then set $z_{k+1} = z_k + \Delta z_k$; otherwise choose any $z_{k+1}$ in the interval (1.59).

**Step 5:** Set $\rho_{k+1} = \rho_k$. If

$$\|g_{k+1} - A^T y_{k+1} - z_{k+1}\| \leq \theta^\text{DP} \mu_k \quad (1.60)$$

and

$$(x_{k+1}, z_{k+1}) \leq n \theta C \mu_k, \quad (1.61)$$

then test whether

$$\xi_{k+1} \leq \theta^\text{FF} \mu_k. \quad (1.62)$$

If all of these inequalities hold, define

$$\mu_{k+1} = \sigma (x_{k+1}, z_{k+1}) / n \quad (1.63)$$

and possibly redefine $\rho_{k+1} \geq \rho_{\min}$, $\omega(\mu_{k+1}) \in (0, \bar{\omega})$.

If (1.60) and (1.61) hold, but (1.62) fails, set $\mu_{k+1} = \mu_k$, and redefine $\rho_{k+1} = \delta \rho_k$ if

$$\Delta v_k = \Delta v^N_k \quad \text{and} \quad \alpha^{(v)}_k \geq \frac{\xi_k}{1 + \xi_k}. \quad (1.64)$$

If either of (1.60) or (1.61) fails, set $\mu_{k+1} = \mu_k$.

In all cases, increment $k$ by one and go back to Step 1.
1.8 Comments on the algorithm

This algorithm suggests a few comments.

1. The requirement that $z_{k+1}$ belongs to the interval (1.59) appears somewhat complex, but it is designed for maximum flexibility in the choice in $z_{k+1}$. The theory below only requires that the components of $z_{k+1}$ are bounded above and away from zero while $\mu_k$ is not updated, and that the choice $z_{k+1} = \mu_k X^{-1} e$ is asymptotically acceptable when $\Delta x_k$ tends to zero. This is similar to the conditions of Gill, Murray, Ponceleon and Saunders (1995), where these bounds are fixed a priori. Note that $z_{k+1} = z_k$ is always a feasible choice when $z_k + \Delta z_k$ does not belong to the interval (1.59), and that then the nonnegativity of $z_{k+1}$ is always guaranteed.

There are many algorithmic possibilities for computing a suitable $z_{k+1}$ when $z_k + \Delta z_k$ does not belong to the interval (1.59). One could, for instance, use a backtracking strategy starting from $z_k + \Delta z_k$, or choose $z_{k+1}$ to minimize $\|X_{k+1} Z_{k+1} e - \mu_k e\|$ subject to being in the desired interval.

Also note that the condition that $z_k + \Delta z_k$ must belong to the interval (1.59) does not restrict the step in $x$.

2. The tests of Step 5 aim to allow for frequent updating of $\mu_k$, and hence for the rapid progress of the algorithm. We will say that iteration $k$ is $\mu$-critical whenever conditions (1.60), (1.61) and (1.62) hold. Condition (1.60) may be viewed as ensuring sufficient dual feasibility (hence $\theta_{DP}$), (1.62) as ensuring sufficient primal feasibility (hence $\theta_{PF}$) and (1.61) as ensuring a sufficient decrease in the value of the complementarity (hence $\theta^c$). This latter condition is inspired by the literature on primal-dual algorithms (see Simantiraki and Shanno (1995), Zhang and Zhang (1994)) or Carpenter, Lustig, Mulvey and Shanno (1993), for instance.

The conditions (1.64) are intended to allow $\rho_k$ to increase when the value of this latter penalty parameter is not large enough to ensure primal feasibility, that is to ensure that the minimum of the merit function lies sufficiently close to the line $\xi = 0$. This is of concern only when a Newton step is used, as the primal-dual step always ensure improved primal feasibility. Hence the first condition. The second guarantees that a significant contribution to the minimization of the merit function is derived from the change in $\xi$.

3. The dependence of the parameters $\omega(\mu_k)$ on $\mu_k$ is introduced with the aim of ensuring that, if $\mu_k$ is decreasing rapidly because of (1.63), the linesearch bound (1.50) should not prevent fast convergence by unduly restricting the stepsize. The threshold $\omega(\mu_k)$ may thus be adapted to avoid this effect. For instance, one might want to choose $\omega(\mu_k)$ to be of the order of $\mu_k$, but the design of a truly efficient strategy will require much more detailed numerical experiments.

4. Suitable values for the constants might be, for instance,

$$\eta = 0.0001, \quad \sigma = \nu_1 = \bar{\omega} = \omega(\mu_k) = 0.01,$$

$$\theta_{DP} = \theta_{PF} = 1, \quad \beta = 0.5, \quad \delta = 10 \quad \theta^c = 99 \quad \text{and} \quad \nu_2 = 100,$$

but this remains to be confirmed by numerical experiences.
5. Observe that the algorithm does not update the value of $y_k$ from iteration to iteration. This is possible because (1.9) and (1.17) directly compute $y_{k+1}^{PD}$ and $y_{k+1}^N$. Thus, although we expect $y_{k+1}$ to converge to the Lagrange multipliers at the solution, these multipliers are recomputed afresh at each iteration.

The fact that $y_k$ is not recurred explicitly has the further advantage that we may modify $\Delta x_k$ when $G_k$ is not second-order sufficient with respect to $A$ without considering any implied change in $y_k$.

6. If primal feasibility is obtained during the course of the calculation, that is if $\xi_k = 0$ for some $k$, the algorithm reduces to a purely (feasible) primal-dual framework.

7. The Newton step $\Delta v_k^N$ can be obtained at low cost from the factorization used to compute $\Delta v_k^{PD}$. Indeed, the system (1.17) is a rank one perturbation of (1.9).

8. As the iterates approach a constrained minimum, we may expect $G_k$ to become second-order necessary with respect to $A$, which implies that no modification of the primal-dual step should be necessary asymptotically, if the threshold value $\lambda$ is chosen small enough. (This is expected because the problem becomes convex in a neighbourhood of such a minimum.) This property would not hold if we had chosen to make $G_k$ positive definite, instead of $N^T G_k N$, possibly resulting in slower asymptotic convergence.

9. Observe that the penalty parameter $\rho_k$ may be updated whenever the barrier parameter $\mu_k$ is reduced. This update may be an increase or a decrease. It provides the possibility of dynamically adapting $\rho_k$ as the algorithm proceeds, without restricting the sequence of penalty parameters to be monotonically increasing.

1.9 Properties of the algorithm

Before proceeding further, we state, for future reference, some useful properties of the algorithm.

**Lemma 3** Let $\{(x_k, \xi_k, z_k)\}$ be a sequence of iterates generated by the algorithm. Then,

(i) the sequence $\{\mu_k\}$ is non-increasing and

$$\mu_{k+1} \neq \mu_k \implies \mu_{k+1} \leq \theta^\epsilon \mu_k,$$

(1.67)

(ii) one has that, for all $k$,

$$\rho_k \geq \rho_{\min}, \quad A \Delta x_k = \Delta \xi_k r_0,$$

(1.68)

and

$$\langle \nabla_v \phi_k, \Delta v_k \rangle \leq -\frac{1}{2} \lambda (\|\Delta v_k\|^2 + (\Delta \xi_k)^2).$$

(1.69)

Furthermore, if $\mu_k = \bar{\mu}$ and $\rho_k = \bar{\rho}$ for some $\bar{\mu}, \bar{\rho} > 0$ and all $k \geq 0$, then there exists a constant $\kappa_1 > 0$ such that

$$0 \leq \xi_k \leq \kappa_1$$

(1.70)

for all $k$. 

12
Proof. The non-increasing nature of the sequence \( \{\mu_k\} \) and (1.67) immediately follow from (1.63), condition (1.61) and the inequality \( \sigma \theta^C < 1 \). The first bound of (1.68) results from the initial value \( \rho_0 \geq \rho_{\min} \) and the fact that \( \rho_k \geq \rho_{\min} \) for all \( k \), because of the mechanism of Step 5. The second equation of (1.68) is a consequence of the mechanism of Steps 2 and 3, (1.43) and the second equation of (1.16). The inequality (1.69) then follows from (1.23), Lemma 2, the first bound of (1.68) and (1.44).

We conclude our proof by showing that, if \( \mu_k \) and \( \rho_k \) are fixed at \( \bar{\mu} \) and \( \bar{\rho} \), respectively, then \( \xi_k \) remains bounded. First notice that the mechanism of Step 2 and Step 3 imposes that, for all \( k \),

\[
\phi(v_k, \bar{\mu}, \bar{\rho}) \leq \phi(v_{k-1}, \bar{\mu}, \bar{\rho}) \leq \phi(v_0, \bar{\mu}, \bar{\rho})
\]

and thus that

\[
(\xi_k + 1)^2 \leq (\xi_0 + 1)^2 + \frac{2}{\rho} [f(x_0) - \bar{\mu}(\log(x_0), e) - (f(x_k) - \bar{\mu}(\log(x_k), e))].
\]

Now, if

\[
(\xi_k + 1)^2 \leq (\xi_0 + 1)^2,
\]

then one obtains that

\[
\xi_k \leq \xi_0.
\]

On the other hand, if (1.73) does not hold, then the expression within brackets in the right-hand side of (1.72) is positive, and thus

\[
(\xi_k + 1)^2 \leq (\xi_0 + 1)^2 + \frac{2}{\rho} [f(x_0) - \bar{\mu}(\log(x_0), e) - \kappa_3],
\]

where \( \kappa_3 = \min_{x \geq 0} [f(x) - \bar{\mu}(\log(x), e)] \) is finite because of (AS2). The bounds (1.74) and (1.75) and the fact that \( \xi_k \geq 0 \) because of (1.53) then yield (1.70), completing the proof \( \square \)

2 Global convergence

2 They are, in fact, the “major” iterations of the algorithm, if expressed as a two-level procedure.
First note that assumption (AS3) implies that there exists a constant $\kappa_8 > 0$ such that, for all $k$,
\[
\|g_k\| \leq \kappa_8 \quad \text{and} \quad \|H_k\| \leq \kappa_8. \tag{2.1}
\]

We next prove a technical result showing under what conditions the primal-dual and Newton steps are bounded when $\mu$ and $\rho$ are fixed.

**Lemma 4** Assume that $\mu_k = \bar{\mu}$ and $\rho_k = \bar{\rho}$ for some $\bar{\mu}, \bar{\rho} > 0$ and all $k \geq 0$. Assume furthermore that there exists a $\kappa_4 > 0$ such that, for all $k$,
\[
\bar{\rho} \leq \kappa_4, \quad \|g_k - \mu_k X_k^{-1} e\| \leq \kappa_4, \quad \text{and} \quad \|G_k\| \leq \kappa_4. \tag{2.2}
\]

Then, there exist a positive constant $\kappa_5 > 0$ such that, for all $k$,
\[
\|\Delta x_k^{PD}\| \leq \kappa_5, \quad \|\Delta \xi_k^{PD}\| \leq \kappa_5 \quad \text{and} \quad \|y_{k+1}^{PD}\| \leq \kappa_5, \tag{2.3}
\]

and
\[
\|\Delta x_k^{N}\| \leq \kappa_5, \quad \|\Delta \xi_k^{N}\| \leq \kappa_5 \quad \text{and} \quad \|y_{k+1}^{N}\| \leq \kappa_5. \tag{2.4}
\]

**Proof.** Consider the primal-dual step first. Writing
\[
\Delta x_k^{PD} = A^T \Delta x_k^{(a)} + N \Delta x_k^{(n)}, \tag{2.5}
\]
we obtain from the second equation of (1.9) that
\[
A A^T \Delta x_k^{(a)} = -r_k \tag{2.6}
\]
which implies, since $A$ has full rank (AS4), that
\[
\|A^T \Delta x_k^{(a)}\| \leq \|A^T (A A^T)^{-1} r_k\| \leq \kappa_1 \|r_0\| \|A\| \|(A A^T)^{-1}\| \overset{\text{def}}{=} \kappa_6, \tag{2.7}
\]
where we have used (1.70) and (1.27) to deduce the last inequality. On the other hand, the first equation of (1.9) gives that
\[
N^T G_k N \Delta x_k^{(n)} = -N^T (g_k - \mu_k X_k^{-1} e) - N^T G_k A^T \Delta x_k^{(a)} \tag{2.8}
\]
The second-order sufficiency of $G_k$ (possibly modified) with respect to $A$, (2.2), (2.7) and (2.8) then ensure that
\[
\lambda \|\Delta x_k^{(n)}\| \leq \kappa_4 (1 + \kappa_6) \|N\|, \tag{2.9}
\]
where $\lambda$ is the smallest eigenvalue of the (possibly modified) $G_k$ restricted to the nullspace of $A$. Combining (2.5), (2.7) and (2.9), we deduce that
\[
\|\Delta x_k^{PD}\| \leq \kappa_6 + \frac{2 \kappa_4}{\lambda} (1 + \kappa_6) \|N\|^2 \overset{\text{def}}{=} \kappa_7. \tag{2.10}
\]
Similarly, we obtain from the first equation of (1.9) that
\[
A A^T y_{k+1}^{PD} = -A (g_k - \mu_k X_k^{-1} e) - A G_k \Delta x_k^{PD}, \tag{2.11}
\]
which yields, using (2.2) and (2.10), that
\[
\|y_{k+1}^{PD}\| \leq \kappa_4 (1 + \kappa_7) \|A\| \|(A A^T)^{-1}\|. \tag{2.12}
\]
Finally, from (1.12) and (1.70),

$$|\Delta \xi_k^{FD}| = |\xi_k| \leq \kappa_1.$$  (2.13)

Together, the bounds (2.10), (2.12) and (2.13) prove (2.3) with

$$\kappa_5 = \kappa_5^{FD} \overset{\text{def}}{=} \max \left[ \kappa_7, \kappa_4 (1 + \kappa_7) \|A\| \| (AAT)^{-1} \|, \kappa_1 \right].$$  (2.14)

Consider now the Newton step. Premultiplying the first equation of (1.17) by $\Delta x_k^N$, we obtain, using successively the second and third equations of the same system, that

$$-(\Delta x_k^N, g_k - \mu_k X_k^{-1} e) = \langle \Delta x_k^N, G_k \Delta x_k^N \rangle + \langle \Delta x_k^N, A^T y_{k+1} \rangle$$

$$= \langle \Delta x_k^N, g_k \Delta x_k^N \rangle + \Delta \xi_k^N (r_0, y_{k+1})$$

$$= \langle \Delta x_k^N, g_k \Delta x_k^N \rangle + \Delta \xi_k^N (\rho_k \Delta \xi_k^N + \rho_k (\xi_k + 1)).$$  (2.15)

Now, using the second-order sufficiency of $G_k$ with respect to $A$ and Lemma 2, we have that

$$\frac{1}{2} \lambda \| \Delta v_k^N \|^2 \leq \langle \Delta x_k^N, G_k \Delta x_k^N \rangle + \rho_k (\Delta \xi_k^N)^2$$

$$= -(\Delta x_k^N, g_k - \mu_k X_k^{-1} e) - \rho_k \Delta \xi_k^N (\xi_k + 1)$$

$$\leq \| \Delta v_k^N \| \| g_k - \mu_k X_k^{-1} e \| + \rho_k (\xi_k + 1).$$  (2.16)

We therefore obtain, using (2.2), (1.70) and the first part of (2.2), that

$$\frac{1}{2} \lambda \| \Delta v_k^N \| \leq \| g_k - \mu_k X_k^{-1} e \| + \rho_k (|\xi_k| + 1) \leq \kappa_4 + \kappa_4 (\kappa_1 + 1).$$  (2.17)

We also obtain from the first equation of (1.17) that

$$AAT y_{k+1}^N = -\left( g_k - \mu_k X_k^{-1} e \right) - A G_k \Delta x_k^N,$$  (2.18)

and thus, using again (2.2), (2.17) and the inequality $\| \Delta x_k^N \| \leq \| \Delta v_k^N \|$, that

$$\| y_{k+1}^N \| \leq \kappa_4 \| A \| \| (AAT)^{-1} \| (1 + \frac{2 \kappa_4}{\lambda} (2 + \kappa_1)).$$  (2.19)

Combining (2.17), (2.19) and $|\Delta \xi_k|^N \leq \| \Delta v_k^N \|$, we obtain (2.4) with

$$\kappa_5 = \kappa_5^N \overset{\text{def}}{=} \max \left[ \frac{2 \kappa_4}{\lambda} (2 + \kappa_1), \kappa_4 \| A \| \| (AAT)^{-1} \| (1 + \frac{2 \kappa_4}{\lambda} (2 + \kappa_1)) \right].$$  (2.20)

The complete result then follows by taking $\kappa_5 = \max[\kappa_5^{PD}, \kappa_5^N]$.

**Lemma 5** Let $\{(x_k, \xi_k, z_k)\}$ be a sequence of iterates generated by the algorithm and assume that

$$\mu_k = \bar{\mu} \text{ and } \rho_k = \bar{\rho}$$  (2.21)

for all $k$. Then, we have that

$$\lim_{k \to \infty} \| \Delta x_k \| = 0,$$  (2.22)

$$\lim_{k \to \infty} \Delta \xi_k = 0,$$  (2.23)

$$\lim_{k \to \infty} X_{k+1} Z_{k+1} e = \bar{\mu} e,$$  (2.24)

$$\lim_{k \to \infty} \| g_{k+1} + A^T y_{k+1} - \bar{\mu} X_{k+1}^{-1} e \| = 0.$$  (2.25)
Proof. We start our proof by noting that, for fixed $\bar{\mu}$ and $\bar{\rho}$, the iteration then reduces to minimizing the function $\phi(v, \bar{\mu}, \bar{\rho})$. Moreover, as a consequence of (2.21), and because the level set

$$L_0 = \{(x, \xi) \in B \times [0, \infty] \mid \phi(x, \xi, \bar{\mu}, \bar{\rho}) \leq \phi(x_0, \xi_0, \bar{\mu}, \bar{\rho})\}\tag{2.26}$$

is bounded away from the boundary of the positive orthant in $x$, we may deduce that, for all $k$,

$$X_k e \geq \kappa_{12} e \tag{2.27}$$

for some $\kappa_{12} \in (0, 1)$. On the other hand, (1.59) and (2.27) imply that

$$\|z_k\| \leq \max \left(\nu_2 \sqrt{n}, \|z_0\|, \nu_2 \frac{\sqrt{n}}{\bar{\mu}}, \nu_2 \frac{\bar{\mu} \sqrt{n}}{\kappa_{12}}\right). \tag{2.28}$$

for all $k$. Combining now this last bound with (2.27) and the second bound of (2.1), we then deduce from the definition of $G_k$ that there exists a $\kappa_{13} > 0$ such that

$$\|G_k\| \leq \kappa_{13}. \tag{2.29}$$

Furthermore, we obtain from (1.46) that we may choose, for each $k$, a $\Delta G_k$ ensuring that $G_k + \Delta G_k$ is second-order sufficient with respect to $A$ (with constant $\lambda$), such that, using (1.46),

$$\|G_k + \Delta G_k\| \leq \|G_k\| + \|\Delta G_k\| \leq (1 + \kappa_2) \kappa_{13} + \lambda \tag{2.30}$$

and the minimum eigenvalue of $G_k + \Delta G_k$ in the nullspace of $A$ is at least $\lambda$. If we now examine the gradient of the merit function with respect to $x$, we verify that

$$\|g_k - \bar{\mu}X_k^{-1} e\| \leq \|g_k\| + \|\bar{\mu}X_k^{-1} e\| \leq \kappa_8 + \bar{\mu} \frac{\sqrt{n}}{\kappa_{12}}, \tag{2.31}$$

where we have used the first bound of (2.1) and (2.27). Combining (2.21), (2.30) and (2.31), we see that all the conditions of Lemma 4 are satisfied for

$$\kappa_4 = \max \left[2 \kappa_{13} + \lambda, \kappa_8 + \bar{\mu} \frac{\sqrt{n}}{\kappa_{12}}, \bar{\rho}\right]. \tag{2.32}$$

We may thus deduce from this lemma that (2.3) and (2.4) hold, which gives that,

$$\|\Delta x_k\| \leq \|\Delta v_k\| \leq \kappa_{14} \tag{2.33}$$

for some $\kappa_{14} > 0$.

We now show that we can deduce a contradiction if the minimization of $\phi(v, \bar{\mu}, \bar{\rho})$ is not successful. To this aim, we make the additional assumption that

$$\|\Delta x_k\| \geq \kappa_{15} \tag{2.34}$$

for all $k \in J$, where $J$ is the index set of a subsequence, and for some $\kappa_{15} \in (0, \kappa_{14})$. We then deduce from (1.69) that, for $k \in J$,

$$\langle \nabla v \phi_k, \Delta v_k \rangle \leq -\frac{1}{2} \lambda \kappa_{15}^2 \overset{\text{def}}{=} -\kappa_{16} < 0. \tag{2.35}$$
We now observe that (1.51), (2.27) and (2.33) give that

$$\hat{\alpha}_k^{(e)} \geq \min \left[ 1, \left(1 - \omega(\mu)\right) \frac{\kappa_{12}}{\kappa_{14}} \right]$$

(2.36)

for all $k$. Furthermore, we note that the mechanism of the algorithm implies that the situation where

$$\alpha_k^{(v)} = \hat{\alpha}_k^{(e)} < 1$$

(2.37)

can only happen for a unique $k$, $\xi_k$ being identically zero (and thus $\hat{\alpha}_k^{(e)}$ being identically one) for all subsequent iterations. Hence, if $k_\xi$ is the index of the iteration where (2.37) holds (defining $k_\xi = \infty$ if (2.37) never holds), we see that

$$\alpha_k^{(v)} < \hat{\alpha}_k^{(e)}$$

if $k < k_\xi$, or

$$\alpha_k^{(v)} \leq 1$$

if $k > k_\xi$.

(2.38)

We therefore conclude that, for $k$ sufficiently large, the inequality $\alpha_k^{(v)} \leq \hat{\alpha}_k^{(e)}$ does not limit the stepsizes in the linesearch procedure (1.56) to a value strictly below one. Moreover, combining (1.54), (2.36) and the definition of $k_\xi$, we have that

$$\hat{\alpha}_k^{(v)} \geq \alpha_k^{(v)} \overset{\text{def}}{=} \min \left[ 1, \left(1 - \omega(\mu)\right) \frac{\kappa_{12}}{\kappa_{14}}, \hat{\alpha}_k^{(e)} \right]$$

(2.39)

for all $k \in J$.

We next consider iteration $k \in J$ and distinguish two cases. The first is when

$$\langle \nabla_v \phi(v_k + \alpha \Delta v_k, \bar{\mu}, \bar{\rho}), \Delta v_k \rangle < \eta \langle \nabla_v \phi_k, \Delta v_k \rangle$$

(2.40)

for all $\alpha \in (0, \alpha^{(v)})$. In the second case, we assume that there exists a (smallest) $\bar{\alpha} \in (0, \alpha^{(v)})$ such that

$$\langle \nabla_v \phi(v_k + \bar{\alpha} \Delta v_k, \bar{\mu}, \bar{\rho}), \Delta v_k \rangle = \eta \langle \nabla_v \phi_k, \Delta v_k \rangle.$$

(2.41)

But (2.35), (2.41) and the mean-value theorem then give that

$$0 < \langle \nabla_v \phi(v_k + \bar{\alpha} \Delta v_k, \bar{\mu}, \bar{\rho}), \Delta v_k \rangle - \langle \nabla_v \phi_k, \Delta v_k \rangle = \bar{\alpha} \langle \nabla^2_v \phi(v_k + \zeta_1 \Delta v_k, \bar{\mu}, \bar{\rho}) \Delta v_k, \Delta v_k \rangle$$

(2.42)

for some $\zeta_1 \in (0, \alpha^{(v)})$. Hence, recalling (2.41), we obtain that

$$\bar{\alpha} = \frac{-\eta \langle \nabla_v \phi_k, \Delta v_k \rangle}{\langle \nabla^2_v \phi(v_k + \zeta_1 \Delta v_k, \bar{\mu}, \bar{\rho}) \Delta v_k, \Delta v_k \rangle}.$$

(2.43)

Observe now that

$$\nabla^2_v \phi(v_k + \zeta_1 \Delta v_k, \bar{\mu}, \bar{\rho}) = \begin{pmatrix}
H(x_k + \zeta_1 \Delta x_k) + \bar{\mu} (X_k + \zeta_1 \Delta X_k)^{-2} & 0 \\
0 & \bar{\rho}
\end{pmatrix}.$$n

(2.44)

Observe also that (AS3), (1.51), (1.54), (1.55) and the fact that $\zeta_1 \leq \alpha^{(v)}$ ensure that

$$\|H(x_k + \zeta_1 \Delta x_k)\| \leq \kappa_{17}$$

(2.45)

for some $\kappa_{17} > 0$. We also deduce from (2.27), (1.51), (1.54), (1.55) and the fact that $\zeta_1 \leq \alpha^{(v)}$ that the components of $x_k + \zeta_1 \Delta x_k$ are bounded below by a positive constant. This fact, (2.45) and (2.44) then imply that

$$\|\nabla^2_v \phi(v_k + \zeta_1 \Delta v_k, \bar{\mu}, \bar{\rho})\| \leq \kappa_{18}$$

(2.46)
for some $\kappa_{18} > 0$. Substituting this bound, (2.33) and (2.35) in (2.43) and using the Cauchy-Schwarz inequality, we obtain that

$$
\overline{\alpha} \geq \frac{(1 - \eta)\kappa_{16}}{\kappa_{18}\kappa_{14}^2}.
$$

(2.47)

Thus, gathering the two cases, we conclude that, for all $k \in J$, (2.40) holds for every $\alpha$ between 0 and

$$
\alpha^* \equiv \min \left( \frac{(1 - \eta)\kappa_{16}}{\kappa_{18}\kappa_{14}^2}, \alpha^{(v)} \right).
$$

(2.48)

Returning to the function $\phi(v, \bar{\mu}, \bar{\rho})$ itself, we therefore obtain that, for each $\alpha \in [0, \alpha^*]$ and all $k \in J$,

$$
\phi(v_k + \alpha\Delta v_k, \bar{\mu}, \bar{\rho}) = \phi_k + \alpha(\nabla_v \phi(v_k + \zeta_2 \Delta v_k, \bar{\mu}, \bar{\rho}), \Delta v_k)
\leq \phi_k + \eta \alpha(\nabla_v \phi_k, \Delta v_k),
$$

(2.49)

where $\zeta_2 \in (0, \alpha)$. As a consequence, the stepsize determined by (1.56) must satisfy

$$
\alpha_k^{(v)} \geq \min \left( \beta \alpha^*, (1 - \omega(\bar{\mu}))(\kappa_{12}/\kappa_{14}) \right) \equiv \kappa_{19}
$$

(2.50)

finally yielding, together with (2.35), that

$$
\phi_{k+1} = \phi(v_k + \alpha_k^{(v)}\Delta v_k, \bar{\mu}, \bar{\rho}) \leq \phi_k - \eta \kappa_{16}\kappa_{19},
$$

(2.51)

for all $k \in J$ sufficiently large. But (2.51) implies that the sequence $\{\phi_k\}$ tends to minus infinity, which is impossible because (AS2) implies that $\phi$ is bounded below in the positive orthant. Hence our assumption (2.34) must be false and we obtain that (2.22) holds. This limit, (2.27) and the inequality $\alpha_k^{(v)} \leq 1$ in turn imply that

$$
\lim_{k \to \infty} \|X_k^{-1} - X_{k+1}^{-1}\| = \lim_{k \to \infty} \max_{i=1, \ldots, n} \left[ \frac{\alpha_k^{(v)}||\Delta x_k||}{\|x_k\|\|x_{k+1}\|} \right] \leq \lim_{k \to \infty} \frac{||\Delta x_k||}{\kappa_{12}} = 0.
$$

(2.52)

But, since

$$
\|z_k + \Delta z_k - \bar{\mu}X_k^{-1}e\| \leq \|z_k + \Delta z_k - \bar{\mu}X_k^{-1}e\| \leq \|X_k^{-1}Z_k\| ||\Delta x_k|| + \bar{\mu}\sqrt{n}\|X_k^{-1} - X_{k+1}^{-1}\|,
$$

(2.53)

where we have used (1.58), we also obtain from (2.22), (2.27), (2.28) and (2.52) that

$$
\lim_{k \to \infty} \|z_k + \Delta z_k - \bar{\mu}X_k^{-1}e\| = 0.
$$

(2.54)

But this limit and the inequalities $\nu_1 < 1$ and $\nu_2 > 1$ give that

$$

\nu_1 \bar{\mu}X_k^{-1}e \leq z_k + \Delta z_k \leq \nu_2 \bar{\mu}X_k^{-1}e
$$

(2.55)

for $k$ sufficiently large. Hence, from the definition of Step 4 of the algorithm, $z_{k+1} = z_k + \Delta z_k$ for sufficiently large $k$. Thus (1.58) yields that

$$
X_{k+1}Z_{k+1}e = X_{k+1}X_k^{-1}(-Z_k\Delta x_k + \bar{\mu}e).
$$

(2.56)

On the other hand, since

$$
\frac{[x_{k+1}]_i}{[x_k]_i} = \frac{[x_k + \alpha_k^{(v)}\Delta x_k]_i}{[x_k]_i} = 1 + \alpha_k^{(v)}[\Delta x_k]_i [x_k]_i,
$$

(2.57)
we deduce from (2.22), (2.27) and \( \alpha_k^{(v)} \in (0, 1] \) that
\[
\lim_{k \to \infty} X_{k+1}X_k^{-1} = I_n, \tag{2.58}
\]
where \( I_n \) is the identity matrix of dimension \( n \). The limit (2.24) then follows from combining (2.56), (2.58), (2.22) and (2.28).

We also note that (1.9), (1.17), (2.27), (2.28) and (2.22) give that
\[
\lim_{k \to \infty} \| y_k + A^Ty_{k+1} - \bar{\mu}X_k^{-1}e \| = 0. \tag{2.59}
\]
We then use the continuity of the gradient, (2.22) and (2.52) to obtain (2.25). Finally, (2.23) follows from (2.22) and the second part of (1.68). \( \square \)

The next stage in our theory is to analyze the situation where the penalty parameter \( \rho_k \) tends to infinity, and show that infeasibilities with respect to the linear equality constraints must then decrease.

**Lemma 6** Let \( \{(x_k, \xi_k, z_k)\} \) be a sequence of iterates generated by the algorithm and define \( I \) to be the index set of all iterations such that \( \rho_k \) is increased at Step 5. Assume furthermore that
\[
\mu_k = \bar{\mu} \tag{2.60}
\]
for all \( k \) and that the subsequence indexed by \( I \) is infinite. Then, there exists an infinite subsequence indexed by \( J \subset I \) such that, for \( k \in J \),
\[
\xi_{k+1} \leq \theta^{PF} \bar{\mu}. \tag{2.61}
\]

**Proof.** Note that the first part of (1.64) implies that \( \Delta v_k = \Delta v_k^S \) for all \( k \in I \). Observe that, for \( k \in I \), some components of \( y_{k+1} \) could be bounded in norm. Let us denote by \( y^B_{k+1} \) the vector whose entries are those of \( y_{k+1} \) in this (possibly empty) components' set and zero elsewhere, and by \( \kappa_{20} > 0 \) the associated upper bound. We thus obtain that, for \( k \in I \),
\[
y_{k+1} = y^H_{k+1} + y^B_{k+1} \quad \text{where} \quad \| y^B_{k+1} \| \leq \kappa_{20}, \tag{2.62}
\]
thus defining \( y^H_{k+1} \).

We now consider two cases. The first is when the set of bounded components of \( y_{k+1} \) is a proper subset of \( \{1, \ldots, m\} \). In this case, there must be an infinite subsequence of \( I \) indexed by \( J \) and some subset \( Z \subset \{1, \ldots, n\} \) such that
\[
\lim_{k \to \infty} \quad \|[A^Ty_{k+1}]_i\| = \lim_{k \in J} \|[A^Ty^H_{k+1}]_i\| = \infty \tag{2.63}
\]
for \( i \in Z \), where we used the fact that \( A \) has full rank (AS4), while
\[
\|[A^Ty_{k+1}]_i\| \leq \kappa_{21} \tag{2.64}
\]
for some \( \kappa_{21} > 0 \) and for \( i \not\in Z \). Now, (1.60) and (2.1) hold for \( k \in J \subset I \), which, with (2.63), implies that
\[
\lim_{k \to \infty} \quad \|[A^Ty_{k+1}]_i\| = \lim_{k \in J} \|[z_{k+1}]_i\| \quad (i \in Z). \tag{2.65}
\]
Since $z_{k+1} > 0$, we immediately deduce from (2.63) and (2.65) that
\begin{equation}
\lim_{k \to \infty} [A^T y_{k+1}]_i = \infty \quad (i \in Z). \tag{2.66}
\end{equation}
Furthermore, (2.65) and (2.66) yield that
\begin{equation}
\lim_{k \to \infty} [z_{k+1}]_i = \infty \quad (i \in Z). \tag{2.67}
\end{equation}
But this last limit is only possible if the last term in the upper bound of (1.59) tends itself to infinity, that is if
\begin{equation}
\lim_{k \to \infty} [x_{k+1}]_i = 0 \quad (i \in Z). \tag{2.68}
\end{equation}
Now, for $k \in J$,
\begin{equation}
\xi_{k+1} \langle r_0, y_{k+1}^\mu \rangle = \langle r_{k+1}, y_{k+1}^\mu \rangle = \langle A(x_{k+1} - x_\odot), y_{k+1}^\mu \rangle = \langle x_{k+1} - x_\odot, A^T y_{k+1}^\mu \rangle, \tag{2.69}
\end{equation}
where $x_\odot$ is given in (AS5) and where we have used (1.27), (1.8) and the identity $Ax_\odot = b$. Clearly, (2.68) gives that
\begin{equation}
\lim_{k \to \infty} [x_{k+1} - x_\odot]_i = -[x_\odot]_i, \tag{2.70}
\end{equation}
for $i \in Z$, which, together with (2.66), (2.64), (2.69) and the inequality $x_\odot > 0$, yields that
\begin{equation}
\lim_{k \to \infty} \xi_{k+1} \langle r_0, y_{k+1}^\mu \rangle = -\infty. \tag{2.71}
\end{equation}
If we now turn to the second case, that is when all components of $y_{k+1}$ are bounded by $\kappa_{20}$, we then have that $y_{k=1}^\mu$ is identically zero and we define $J = I$.
In both cases, we obtain from the Cauchy-Schwarz inequality and $\xi_{k+1} \geq 0$ that
\begin{equation}
\langle r_0, y_{k+1} \rangle = \langle r_0, y_{k+1}^B \rangle + \langle r_0, y_{k+1}^\mu \rangle \leq \kappa_{20} \|r_0\| \tag{2.72}
\end{equation}
for $k \in J$ sufficiently large, where we used (2.71) to obtain the last inequality if $y_{k+1}^\mu$ is nonzero. Now the third equation of (1.16), the bound $\alpha_k^{(v)} \leq 1$ and the second part of (1.64) then give that, for $k \in J$,
\begin{equation}
\xi_{k+1} = \xi_k + \alpha_k^{(v)} \Delta \xi_k = \frac{\alpha_k^{(v)}}{\rho_k} \langle r_0, y_{k+1} \rangle + \xi_k - \alpha_k^{(v)} (\xi_k + 1) \leq \frac{1}{\rho_k} \langle r_0, y_{k+1} \rangle. \tag{2.73}
\end{equation}
Substituting (2.72) in (2.73) then gives that, for $k \in J$ sufficiently large,
\begin{equation}
\xi_{k+1} \leq \frac{\kappa_{20} \|r_0\|}{\rho_k}. \tag{2.74}
\end{equation}
This and the fact that $\rho_k$ tends to infinity ensures that (2.61) holds for $k \in J$ sufficiently large, as required. □

We are now ready to prove our main convergence result.
Theorem 7 Let \( \{x_k, \xi_k, z_k\} \) be a sequence of iterates generated by the algorithm and define

\[
\mathcal{K} = \{ k \mid \mu_k < \mu_{k-1} \}. \tag{2.75}
\]

Then, \( \mathcal{K} \) is infinite and we have that

\[
\lim_{k \to \infty, k \in \mathcal{K}} X_k Z_k = 0, \tag{2.76}
\]

\[
\lim_{k \to \infty, k \in \mathcal{K}} \| g_k + A^T y_k - z_k \| = 0 \tag{2.77}
\]

and

\[
\lim_{k \to \infty, k \in \mathcal{K}} \| Ax_k - b \| = 0. \tag{2.78}
\]

Proof. In order to prove this convergence result, we will now consider the behaviour of the algorithm if convergence never occurs, and later deduce that this behaviour is impossible. Assume therefore, for the purpose of establishing a contradiction, that, for all \( k \),

\[
\mu_k \geq \bar{\mu} > 0. \tag{2.79}
\]

Because Lemma 3 ensures that the sequence \( \{\mu_k\} \) is non-increasing, (2.79) implies that the update (1.63) is never performed for \( k \) sufficiently large. and we may thus assume, without loss of generality, that

\[
\mu_k = \bar{\mu} \tag{2.80}
\]

for all \( k \geq 0 \).

Assume first that

\[
\Delta v_k = \Delta v_k^N \tag{2.81}
\]

and

\[
\rho_k = \bar{\rho} \tag{2.82}
\]

hold for all \( k \) sufficiently large. Because of equalities (2.80) and (2.82), we may then apply Lemma 5 and deduce that (2.22), (2.23), (2.24) and (2.25) hold. But these limits imply that conditions (1.60) and (1.61) are satisfied for \( k \) sufficiently large. Furthermore (1.51), (2.22) and (2.27) ensure that \( \alpha_k^{(x)} = 1 \) for all \( k \) sufficiently large. Moreover, as (2.80) guarantees that (1.62) cannot be true, (1.53) ensures that \( \alpha_k^{(x)} = 1 \) for all \( k \) sufficiently large. Hence \( \alpha_k^{(v)} = 1 \), and (1.64) are satisfied for all \( k \) sufficiently large. Since \( \rho_k \) remains constant, the mechanism of Step 5 then ensures that (1.62) must also be satisfied for such \( k \). As a consequence, \( \mu_k \) is eventually reduced according to (1.63), which contradicts (2.80). Hence, if \( \mu_k \) remains constant and (2.81) holds for all sufficiently large \( k \), \( \rho_k \) must tend to infinity and is increased in Step 5, for some infinite subsequence \( I \). We may then apply Lemma 6 and deduce (2.61) for some subsequence \( J \) for which conditions (1.60), (1.61) and (1.62) hold. As above, this in turn implies that \( \mu_k \) is reduced according to (1.63), again contradicting (2.80). We therefore deduce that (2.81) cannot hold for all sufficiently large \( k \) if (2.80) holds. As a consequence, if this last relation holds, there must exist an infinite subsequence indexed by \( L \) such that

\[
\Delta v_k = \Delta v_k^{PD} \tag{2.83}
\]
for \( k \in L \). Applying now Lemma 6 as above, we also conclude that, if \( \rho_k \) is increased infinitely often in Step 5, then \( \mu_k \) must be reduced, which is impossible because of (2.80). As a consequence, we therefore deduce that \( \rho_k \) remains constant (and equal to some \( \bar{\rho} \)) for sufficiently large \( k \). We may then apply Lemma 5 again, and deduce that (2.22), (2.24) and (2.25) hold for sufficiently large \( k \). But the first of these limits, the third part of (1.68) and the second block of (1.9) together then imply that

\[
\| r_0 \| \lim_{k \to \infty} \xi_k = \lim_{k \to \infty} \| r_k \| \leq \| A \| \lim_{k \to \infty} \| \Delta z_k^P \| = 0 \tag{2.84}
\]

for \( k \in L \). Once more, we see that, for \( k \in L \) sufficiently large, \( \mu_k \) must then be reduced using (1.63), since (2.24), (2.25) and (2.84) ensure that a \( \mu \)-critical iteration must occur eventually. This again contradicts (2.80), finally proving that this last assumption is impossible.

Hence \( \mu_k \) is not bounded away from zero. But (1.63) implies that \( \mu_k > 0 \) for all \( k \), and thus that the subsequence indexed by \( \mathcal{K} \) is infinite, and we deduce from condition (1.67) of Lemma 3 and the inequality \( \sigma \theta^c < 1 \) that

\[
\lim_{k \to \infty} \mu_k = 0. \tag{2.85}
\]

Recalling now the definition of \( \mathcal{K} \), the index set of all iterations immediately following an update of \( \mu_k \) using (1.63), we then see that (1.61) implies that

\[
X_k Z_k e \leq \langle x_k, z_k \rangle e = \frac{n}{\sigma} \mu_k e \tag{2.86}
\]

for \( k \in \mathcal{K} \). But this inequality and (2.85) together yield the limit (2.76). Combining (2.76), (1.63), (1.60) and (1.62), one obtains (2.77) and (2.78), which concludes the proof. □

3 Algorithmic variants and further comments

3.1 A more general monotonic update for the penalty parameter

The link of the penalty parameter to the current average complementary slackness value (as given by (1.63)) can also be relaxed somewhat. Indeed, the only formal role of this choice for \( \mu_k \) is to force optimality when \( \mu_k \) tends to zero using (1.60) and to ensure that \( \mu_k \) is decreased when it is updated. We can consider a variant of our algorithm where the distinct values of the sequence \( \{\mu_k\} \) are chosen a priori as a sequence \( \{\hat{\mu}_k\} \) converging to zero. In this case, Step 5 has the following form.

**Step 5:** Set \( \rho_{k+1} = \rho_k \). If the conditions (1.60) and (1.61) hold, then test if (1.62) also hold. If this is the case, then decrease the penalty parameter by setting \( \mu_{k+1} \) to the value immediately following \( \mu_k \) in the sequence \( \{\hat{\mu}_k\} \); otherwise set \( \mu_{k+1} = \mu_k \) and reset \( \rho_k = 0 \) if both conditions (1.64) are satisfied.

In all cases, increment \( k \) by one and go back to Step 1.

Whether choosing an a priori sequence \( \{\hat{\mu}_k\} \) is better than determining the sequence of penalty parameter using (1.63) remains to be seen in practice. One could of course argue that an a priori subsequence leaves more freedom to the user, but any a priori choice is also somewhat arbitrary and may not reflect what is actually happening in the course of the calculation. Leaving decisions
to the users may also not be so desirable from their own point of view. The advantage of (1.63) is that it depends on the current values of the variables and may therefore set more realistic goals for the minimization of the merit function. The main reason to mention the variant discussed in this paragraph is that it makes our algorithm similar to the framework of Gill et al. (1995).

3.2 Introducing non-monotonicity

The monotonic character of the sequence of \( \{\phi_k\} \) for fixed \( \mu_k \) and \( \rho_k \), and that of the sequence \( \{\mu_k\} \) itself are not necessary. For the values of the merit function, one could think of modifying the linesearch (1.56), as in Grippo, Lampariello and Lucidi (1986) or Toint (1996), resetting the process when \( \mu_k \) is updated. One could also relax the first part of condition (1.61) to allow a non-monotone behaviour of the penalty parameter, replacing it by the condition

\[
(\langle z_{k+1}, z_{k+1}\rangle) \leq n^{\gamma^C} \max_{j=0, \ldots, p} \mu_{k-j} \tag{3.1}
\]

for some integer \( p > 0 \). More sophisticated schemes (see Toint, 1996) are possible if this type of relaxation appears to be useful in practice.

It may also be advisable from a practical point of view to considerably relax the conditions of Step 3 for the first few iterations, in order to let the algorithm choose a suitable value of the penalty parameter, which may result in better overall performance. Again, this has to be verified in numerical experiments.

3.3 Further updating the penalty parameter \( \rho_k \)

If the primal-dual step does not give a sufficient descent, that is if (1.44) does not hold in Step 2, one might, in view of (1.25), think of simply increasing the penalty parameter \( \rho_k \) at this stage. Although such an increase can be accepted finitely often (for each value of \( \mu_k \)), there is a danger that negative interaction between the barrier and the penalty term might require an infinite sequence of such increases, which would result in very poor numerical behaviour and also ruin the convergence theory presented above. This would happen if the algorithm gives too much weight to primal feasibility when the logarithmic singularity is active, inducing a large gradient \( \nabla z \phi_k \), resulting in an undesirable loop where the iterates approach the boundary of the positive orthant and \( \rho_k \) tends to infinity.

Thus, increasing \( \rho_k \) is Step 2 may be accepted, but should be monitored to avoid this difficulty. As already indicated, the simplest strategy is to only allow a finite increase in \( \rho_k \) as long as the barrier parameter is not updated. Other more elaborate strategies may result from continued numerical experience with the algorithm.

3.4 A more general update for the dual variables

The last variant that we consider consists of relaxing even further the conditions of Step 4 on the update of the dual variables \( z \). In the algorithm described above, we have enforced the choice \( z_{k+1} = z_k + \Delta z_k \) whenever this vector falls in the interval (1.59). This can be relaxed somewhat, in that our theory still holds if we only require that \( z_{k+1} \) is any vector satisfying the bounds given by (1.59) with the property that

\[
\lim_{k \to \infty} \Delta z_k = 0 \text{ implies } \lim_{k \to \infty} X_{k+1} Z_{k+1} = \mu_k e. \tag{3.2}
\]
This implication is indeed all we need to obtain the limit (2.24) from (2.22) at the end of the proof of Lemma 5.

The main interest of this slight extension is that it now covers the case where

$$z_k = \mu_k X_k^{-1} e.$$  \hspace{1cm} (3.3)

If the choice (3.3) is made, the algorithm reduces, for iteration $k$, to a pure primal method in that $z_k$ is entirely eliminated from the computation: $\Delta z_k$ need not be computed and (1.58) may thus be skipped altogether. We then obtain that

$$G_k = H_k + \mu_k X_k^{-2},$$  \hspace{1cm} (3.4)

which is exactly the Hessian of the merit function $\phi(v_k, \mu_k, \rho_k)$ in the $x$-space. This may be attractive if one wishes to exploit directions of negative curvature for the merit function, as they then correspond to linear combinations of eigenvectors of $G_k$ associated with negative eigenvalues. Again, the detail of these considerations is beyond the scope of the present paper and we postpone their presentation for future work.

4 Preliminary numerical tests

In order to investigate the effectiveness of the method discussed in this paper, we have written a prototype fortran 90 implementation of the algorithm proposed in Section 1.7 to solve quadratic programs, that is for problems for which $f(x)$ is a quadratic function. In this implementation, $z_{k+1}$ is simply chosen as $z_k + \alpha_k^{(z)} \Delta z_k$, where $\alpha_k^{(z)}$ is the minimum of 1 and the largest stepsize such that $z_k + \alpha_k^{(z)} \Delta z_k$ remains in the interval (1.59).

The solution of linear systems of the form

$$\begin{pmatrix} G & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ y \end{pmatrix} = - \begin{pmatrix} g \\ r \end{pmatrix}$$

lie at the heart of the algorithm. For convenience, we have used the Harwell Subroutine Library (1995) package MA27 to solve such systems. The multifrontal scheme used (see, Duff and Reid, 1983) has the additional advantages of being able to cope with large, sparse systems and of reporting the inertia of the relevant coefficient matrix, $K$. Although some authors (for instance, Gill, Murray, Saunders and Wright, 1990) have reported that such an approach is handicapped by the severe indefiniteness of $K$, we followed the advice of Gill, Murray, Poncélion and Saunders (1991) and use a very small pivot threshold ($10^{-6}$) together with iterative refinement as an effective means of solution. If MA27 reports that $G$ is not second-order sufficient, we use the naive expedient of replacing $G$ by $G + ||G|| I$. More sophisticated strategies are being considered (see Gould, 1995), but are beyond the scope of this paper.

The actual algorithm implemented is the obvious generalization of the algorithm described above designed to cope with simple bounds of the form $l \leq x \leq u$ rather than nonnegativities alone. All fixed variables are removed automatically and the minimization performed with respect to the remaining variables. A given starting point $x$ is adjusted so that each component lies at least a distance ten on the feasible side of its nearest bound; if this is impossible the mid point between the two bounds is chosen. Similarly, the dual variable associated with each simple bound
is supplied by the user (we used zero in our tests) and adjusted so that it is at least a distance 10 to the feasible side of its relevant dual bound. We use \( \mu_0 = \langle x_0, z_0 \rangle / n \) and the parameter values suggested in (1.65)(1.66). The algorithm is halted as soon as the norm of the residual of (1.2) is smaller than \( 10^{-4} \), or more than 1000 iterations have been performed.

To test our algorithm, we have selected all of the larger quadratic programs in the CUTE test set (see, Bongartz et al., 1995). Although it is desirable in practice to preprocess the problems (for instance, to remove redundant constraints and scale the problem, see Andersen, Gondzio, Mézáros and Xu, 1996), we have not done so.

<table>
<thead>
<tr>
<th>Name</th>
<th>n</th>
<th>m</th>
<th>type</th>
<th>its</th>
<th>Nwtn</th>
<th>mods</th>
<th>time</th>
<th>VE09-its</th>
<th>VE09-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUG2DCQP</td>
<td>3280</td>
<td>1600</td>
<td>C</td>
<td>21</td>
<td>5</td>
<td>0</td>
<td>4.86</td>
<td>3112</td>
<td>133.15</td>
</tr>
<tr>
<td>AUG2DQP</td>
<td>3280</td>
<td>1600</td>
<td>C</td>
<td>21</td>
<td>3</td>
<td>0</td>
<td>4.72</td>
<td>3019</td>
<td>127.71</td>
</tr>
<tr>
<td>AUG3DCQP</td>
<td>3873</td>
<td>1000</td>
<td>C</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>5.00</td>
<td>3056</td>
<td>106.58</td>
</tr>
<tr>
<td>AUG3DQP</td>
<td>3873</td>
<td>1000</td>
<td>C</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>6.26</td>
<td>2097</td>
<td>71.44</td>
</tr>
<tr>
<td>BLOCKQP1</td>
<td>2006</td>
<td>1001</td>
<td>NC</td>
<td>26</td>
<td>0</td>
<td>18</td>
<td>6.34</td>
<td>1006</td>
<td>28.76</td>
</tr>
<tr>
<td>BLOCKQP2</td>
<td>2006</td>
<td>1001</td>
<td>NC</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>2.92</td>
<td>1006</td>
<td>40.42</td>
</tr>
<tr>
<td>BLOCKQP3</td>
<td>2006</td>
<td>1001</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>1.60</td>
<td>1597</td>
<td>68.96</td>
</tr>
<tr>
<td>BLOWEYA</td>
<td>2002</td>
<td>1002</td>
<td>C</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1.29</td>
<td>1497</td>
<td>67.86</td>
</tr>
<tr>
<td>BLOWEYB</td>
<td>2002</td>
<td>1002</td>
<td>C</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1.72</td>
<td>1697</td>
<td>53.97</td>
</tr>
<tr>
<td>BLOWEYC</td>
<td>2002</td>
<td>1002</td>
<td>C</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1.72</td>
<td>1697</td>
<td>53.97</td>
</tr>
<tr>
<td>CVXQP1</td>
<td>1000</td>
<td>500</td>
<td>C</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>44.44</td>
<td>861</td>
<td>70.92</td>
</tr>
<tr>
<td>CVXQP2</td>
<td>1000</td>
<td>250</td>
<td>C</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>19.15</td>
<td>370</td>
<td>13.50</td>
</tr>
<tr>
<td>CVXQP3</td>
<td>1000</td>
<td>750</td>
<td>C</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>132.40</td>
<td>1389</td>
<td>107.11</td>
</tr>
<tr>
<td>DUALC1</td>
<td>223</td>
<td>215</td>
<td>C</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>1.17</td>
<td>12</td>
<td>0.23</td>
</tr>
<tr>
<td>DUALC2</td>
<td>235</td>
<td>229</td>
<td>C</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>0.83</td>
<td>14</td>
<td>0.25</td>
</tr>
<tr>
<td>DUALC5</td>
<td>285</td>
<td>278</td>
<td>C</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0.37</td>
<td>10</td>
<td>0.37</td>
</tr>
<tr>
<td>DUALC8</td>
<td>510</td>
<td>503</td>
<td>C</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1.05</td>
<td>11</td>
<td>0.94</td>
</tr>
<tr>
<td>GOULDQP2</td>
<td>699</td>
<td>349</td>
<td>C</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
<td>251</td>
<td>1.84</td>
</tr>
<tr>
<td>GOULDQP3</td>
<td>699</td>
<td>349</td>
<td>C</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.35</td>
<td>463</td>
<td>2.52</td>
</tr>
<tr>
<td>KSP</td>
<td>1021</td>
<td>1001</td>
<td>C</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>6.38</td>
<td>1388</td>
<td>36.42</td>
</tr>
<tr>
<td>MOSARQP1</td>
<td>1500</td>
<td>600</td>
<td>C</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>2.01</td>
<td>5859</td>
<td>91.63</td>
</tr>
<tr>
<td>MOSARQP2</td>
<td>1500</td>
<td>600</td>
<td>C</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>1.60</td>
<td>1679</td>
<td>27.73</td>
</tr>
<tr>
<td>NCVXQP1</td>
<td>1000</td>
<td>500</td>
<td>NC</td>
<td>956</td>
<td>297</td>
<td>927</td>
<td>3762.58</td>
<td>1561</td>
<td>51.04</td>
</tr>
<tr>
<td>NCVXQP2</td>
<td>1000</td>
<td>500</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>1840</td>
<td>61.50</td>
</tr>
<tr>
<td>NCVXQP3</td>
<td>1000</td>
<td>500</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>470</td>
<td>1913.23</td>
<td>too ill-cond. basis</td>
</tr>
<tr>
<td>NCVXQP4</td>
<td>1000</td>
<td>250</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>470</td>
<td>649</td>
<td>2.97</td>
</tr>
<tr>
<td>NCVXQP5</td>
<td>1000</td>
<td>250</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>319</td>
<td>565</td>
<td>2.78</td>
</tr>
<tr>
<td>NCVXQP6</td>
<td>1000</td>
<td>750</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>319</td>
<td>532</td>
<td>3.53</td>
</tr>
<tr>
<td>NCVXQP7</td>
<td>1000</td>
<td>750</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>319</td>
<td>1901</td>
<td>141.70</td>
</tr>
<tr>
<td>NCVXQP8</td>
<td>1000</td>
<td>750</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>0</td>
<td>0</td>
<td>319</td>
<td>1567</td>
<td>120.56</td>
</tr>
</tbody>
</table>

Table 1: Preliminary numerical results (1)

In Tables 1 and 2, we give the results of our preliminary tests. They were performed in double precision on an IBM RISC System/6000 3B7 workstation with 64 Megabytes of RAM, using the xlf90 compiler and optimization level -O3. For each example, we report its name along with its dimensions (n is the number of variables, m the number of constraints), the problem
type (C for convex, SOS for second-order sufficient and NC for non-convex and not second-order sufficient), the number of iterations performed (its), the number of these which were Newton (1.17) iterations (Nwtm) and the number for which $G$ was modified (mods), and the time taken in seconds (time). For comparison, the tables also show the number of iterations and time taken by a fortran-90 version of VE09, a quadratic programming subroutine from the Harwell Subroutine Library (VE09-its and VE09-time, respectively). This latter algorithm is designed to handle non-convex problems and is of the active-set type, each of its iterations corresponding to a pivoting operation. The reader is referred to Gould (1991) for further details on this method. We also ran tests using MINOS of Murtagh and Saunders (1993) which we do not report here because they are quantitatively similar to those obtained with VE09.

<table>
<thead>
<tr>
<th>Name</th>
<th>n</th>
<th>m</th>
<th>type</th>
<th>its</th>
<th>Nwtm</th>
<th>mods</th>
<th>time</th>
<th>VE09-its</th>
<th>VE09-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRIMALC1</td>
<td>239</td>
<td>9</td>
<td>C</td>
<td>83</td>
<td>40</td>
<td>0</td>
<td>2.62</td>
<td>20</td>
<td>0.17</td>
</tr>
<tr>
<td>PRIMALC2</td>
<td>238</td>
<td>7</td>
<td>C</td>
<td>61</td>
<td>4</td>
<td>0</td>
<td>1.69</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>PRIMALC5</td>
<td>295</td>
<td>8</td>
<td>C</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0.43</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>PRIMALC8</td>
<td>528</td>
<td>8</td>
<td>C</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0.76</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>PRIMAL1</td>
<td>410</td>
<td>85</td>
<td>C</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>3.13</td>
<td>361</td>
<td>4.12</td>
</tr>
<tr>
<td>PRIMAL2</td>
<td>745</td>
<td>96</td>
<td>C</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>3.02</td>
<td>677</td>
<td>12.25</td>
</tr>
<tr>
<td>PRIMAL3</td>
<td>856</td>
<td>111</td>
<td>C</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>14.22</td>
<td>798</td>
<td>35.49</td>
</tr>
<tr>
<td>PRIMAL4</td>
<td>1564</td>
<td>75</td>
<td>C</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>6.51</td>
<td>1515</td>
<td>40.16</td>
</tr>
<tr>
<td>QPCBOE11</td>
<td>726</td>
<td>351</td>
<td>C</td>
<td>113</td>
<td>11</td>
<td>0</td>
<td>8.45</td>
<td>823</td>
<td>6.41</td>
</tr>
<tr>
<td>QPCBOE12</td>
<td>305</td>
<td>166</td>
<td>C</td>
<td>109</td>
<td>4</td>
<td>0</td>
<td>3.38</td>
<td>303</td>
<td>1.14</td>
</tr>
<tr>
<td>QPCSTAIR</td>
<td>614</td>
<td>356</td>
<td>C</td>
<td>174</td>
<td>8</td>
<td>0</td>
<td>13.26</td>
<td>987</td>
<td>16.05</td>
</tr>
<tr>
<td>QPNBOE11</td>
<td>726</td>
<td>351</td>
<td>NC</td>
<td>&gt; 1000</td>
<td>652</td>
<td>1</td>
<td>639</td>
<td>61.88</td>
<td>736</td>
</tr>
<tr>
<td>QPNBOE12</td>
<td>305</td>
<td>166</td>
<td>NC</td>
<td>652</td>
<td>1</td>
<td>639</td>
<td>61.88</td>
<td>299</td>
<td>1.14</td>
</tr>
<tr>
<td>QPNSTAIR</td>
<td>614</td>
<td>356</td>
<td>NC</td>
<td>226</td>
<td>13</td>
<td>207</td>
<td>30.39</td>
<td>993</td>
<td>15.61</td>
</tr>
<tr>
<td>SOSQP1</td>
<td>2000</td>
<td>1001</td>
<td>SOS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.73</td>
<td>996</td>
<td>14.49</td>
</tr>
<tr>
<td>STCQP1</td>
<td>4097</td>
<td>2052</td>
<td>NC</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2845</td>
<td>67.81</td>
</tr>
<tr>
<td>STCQP2</td>
<td>4097</td>
<td>2052</td>
<td>NC</td>
<td>11</td>
<td>0</td>
<td>4</td>
<td>51.09</td>
<td>2040</td>
<td>98.22</td>
</tr>
<tr>
<td>STNQP1</td>
<td>4097</td>
<td>2052</td>
<td>NC</td>
<td>11</td>
<td>0</td>
<td>4</td>
<td>51.09</td>
<td>3158</td>
<td>68.01</td>
</tr>
<tr>
<td>STNQP2</td>
<td>4097</td>
<td>2052</td>
<td>NC</td>
<td>24</td>
<td>0</td>
<td>15</td>
<td>201.12</td>
<td>1408</td>
<td>39.11</td>
</tr>
<tr>
<td>UBH1</td>
<td>909</td>
<td>600</td>
<td>C</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0.29</td>
<td>315</td>
<td>5.12</td>
</tr>
<tr>
<td>YAO</td>
<td>1002</td>
<td>500</td>
<td>C</td>
<td>847</td>
<td>7</td>
<td>0</td>
<td>38.34</td>
<td>3</td>
<td>2.06</td>
</tr>
</tbody>
</table>

Table 2: Preliminary numerical results (2)

We immediately note that the primal-dual algorithm performs well on convex problems (C and SOS), with the possible exception of YAO. On the other hand, its performance on the non-convex (NC) ones is somewhat disappointing. A closer examination of these runs indicates that our naive matrix modification technique is really too naive; when the Hessian involves many negative eigenvalues, these appear to be removed one at a time, resulting in a large number of iterations before second-order sufficiency is achieved. A more sophisticated way of treating negative curvature directions is therefore highly desirable. Generally however, given the crude nature of the present preliminary implementation, the new primal-dual method definitely shows some potential.
5 Conclusion

We have presented a primal-dual algorithmic framework whose merit function is adapted to problems with non-convex objective functions. We also proved global convergence for this framework under standard assumptions. Finally some preliminary numerical results have been given and discussed, indicating a clear potential for further research.

In particular, the use of negative curvature directions appears to require more sophistication. Although the current method works in its current naive form, it converges slowly for problems involving massive indefiniteness. Less naive strategies are thus needed and are the object of current investigations. Other potentially promising developments have been outlined in Section 3.

Unsurprisingly there are disadvantages to the approach we have taken. The primary numerical linear algebraic computation is essentially a calculation involving the Karush-Kuhn-Tucker matrix

\[
\begin{pmatrix}
H_k + X_k^{-1}Z_k & A^T \\
A & 0
\end{pmatrix}
\]  

(5.6)

which is inherently non-trivial to handle because one expects small components in \(z_k\) without corresponding small components in \(z_k\). The analogous matrix in the case of linear programming is the matrix

\[
\begin{pmatrix}
I + X_k^{-1}Z_k & A^T \\
A & 0
\end{pmatrix}
\]  

(5.7)

The fact that in this case the upper left-hand block is diagonal (and for non-degenerate problems this matrix is asymptotically non-singular) makes this form of ill-conditioning easier to handle, (see for example, Wright (1992)). However, Ponceleón (1990) and Forsgren, Gill and Shinnerl (1996) show how one can treat the general case.

A more direct concern is that it is inappropriate to use the normal equations when considering (5.6) instead of (5.7). Many authors have suggested using a direct factorization of (5.6)/(5.7) (see for example Duff, Gould, Reid, Scott and Turner (1991), Fourer and Mehrotra (1993), Vanderbei and Carpenter (1993) and Andersen et al. (1996)) which can be very successful. Other issues we would like to consider in future include trying to justify why a primal-dual approach should be more successful globally even for nonconvex problems than a primal approach, and trying to explain why the central path appears to be so important. Since it is also generally recommended that, at least in the case of interior point approaches to the linear programming problem, one makes use of predictor-corrector techniques to enhance performance, we remark that we wish to extend the methods considered here to include such improvements.

6 Acknowledgements

The authors are pleased to acknowledge the useful comments made by M. H. Wright.
References


