The Convergence of an Interior-Point Method Using Modified Search Directions in Final Iterations

WEICHIUNG WANG
Department of Mathematics Education
National Tainan Teachers College Tainan 700, Taiwan, R.O.C.
wwang@ipx.ntntc.edu.tw

(Received September 2000; revised and accepted October 2001)

Abstract—we provide an asymptotic analysis of a primal-dual algorithm for linear programming that uses modified search directions in the final iterations. The algorithm determines the search directions by solving the normal equations using the preconditioned conjugate gradient algorithm. Small dual slack variables are slightly perturbed in the later stage of the interior-point algorithm to obtain better conditioned systems without interfering with convergence. The modification and its motivation are discussed, and a convergence analysis of the resulting algorithm is presented. The analysis shows the iterates of the modified system converge to the solution of the Karush-Kuhn-Tucker optimality system associated with the Lagrangian of the logarithmic barrier subproblem. The global convergence of the interior-point method is thus established. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords—Linear programming, Interior-point methods, Modified search directions, Preconditioned conjugate gradient, Normal equations, Convergence analysis.

1. INTRODUCTION

Interior-point methods [1-3] have been widely used to solve linear programming problems arising in many applications. A standard linear programming problem can be written as

\[ \min c^T x, \]
\[ \text{s.t. } Ax = b, \]
\[ x \geq 0, \]

where \( c \) and \( x \) are real \( n \)-vectors, \( b \) is a real \( m \)-vector, and \( A \in \mathbb{R}^{m \times n} \) is a real matrix of rank \( m \) with \( m < n \). There are a variety of efficient implementations of interior-point algorithms, especially for large scale problems (e.g., [4-6]).

Typical interior-point methods involve both outer and inner loops. Each iteration in the outer loop is associated with a positive barrier parameter \( \mu \) taken from a decreasing sequence \( \{\mu_k\} \) that...
converges to zero. A logarithmic barrier function is applied to the inequality in (1) to obtain the barrier subproblem associated with the standard linear programming problem

$$\min \quad c^T x - \mu \sum_{i=1}^{n} \ln \chi_i,$$

s.t. \quad Ax = b, \quad (2)

where \( x = (x_1, x_2, \ldots, x_n)^T \). The equality constraint of the barrier subproblem is then removed to obtain the Lagrangian function

$$c^T x - \mu \sum_{i=1}^{n} \ln \chi_i + y^T (Ax - b),$$

where the \( m \)-vector \( y \) contains the Lagrange multipliers. In the inner loop, interior-point methods find a stationary point of the Lagrangian using Newton's method applied to the first optimality conditions associated with the Lagrangian

$$XZe - \mu e = 0,$$
$$c - A^T y - z = 0,$$
$$Ax - b = 0. \quad (4)$$

The \( n \)-vector \( z \) is a dual slack variable and \( e \) is a \( n \times 1 \) vector with all 1s. \( X \) and \( Z \) are diagonal matrices containing the entries of \( x \) and \( z \) (respectively) on their main diagonals. The iterates obtained by solving these problems converge to an optimal solution to the linear programming problem.

The most computationally intensive part of interior-point methods is determining the Newton search direction for (4). This direction is usually determined by solving the Karush-Kuhn-Tucker (KKT) system

$$\begin{pmatrix} X & Z & 0 \\ I & 0 & A^T \\ 0 & A & 0 \end{pmatrix} \begin{pmatrix} \Delta z \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} Xz - \mu e \\ A^T y + z - c \\ Ax - b \end{pmatrix} \quad (5)$$

or alternatively, the normal equations

$$(A\Theta A^T) \Delta y = A\Theta (r_d + Ze - \mu X^{-1}e) + r_p, \quad (6)$$

where \( r_p = b - Ax, \ r_d = A^T y + z - c, \) and \( \Theta = Z^{-1}X \). If the normal equations are solved for \( \Delta y \), then \( \Delta x \) and \( \Delta z \) can be computed by

$$\Delta x = \Theta (A^T \Delta y - r_d - Ze + \mu X^{-1}e) \quad (7)$$

and

$$\Delta z = r_d - A^T \Delta y. \quad (8)$$

Direct methods that rely on sparse matrix factorizations have been the most popular approaches for solving the normal equations (e.g., [2,7,8]). These methods take advantage of the fact that the sparsity pattern of the matrix \( A\Theta A^T \) remains unchanged in the course of interior-point methods. In contrast, iterative methods are often used for solving the KKT system (e.g., [9,10]) and sometimes used for solving the normal equations (e.g., [11-16]). Iterative methods can be stopped adaptively whenever accuracy requirements have been satisfied. An additional advantage is that they do not need the KKT matrix or the matrix \( A\Theta A^T \) to be formed explicitly. However, it is a challenge to find effective preconditioners when the \( D \) matrix is (rapidly) changing and to determine the stopping criteria over the different stages of the interior-point methods.
In order to efficiently solve the normal equations by preconditioned conjugate gradients, Wang and O'Leary [17] proposed adaptive preconditioning strategies. They use either a Cholesky factorization or a sequence of rank-1 updates to determine preconditioners. Their algorithm, however, switches to a direct method in the final iterations of the interior-point methods due to ill-conditioning of \( \Theta \). To improve the adaptive algorithm, Wang [18] implemented an algorithm that uses modified search directions in the later stage. The algorithm perturbs large entries of the diagonal matrix \( \Theta \) in the left-hand side of equation (6) to speed up the convergence of preconditioned conjugate gradients.

While [18] focused on implementation details and presented promising numerical results, here we give detailed convergence analysis of the modified search direction algorithm. Assuming the dual solution set is bounded and that each step for the \( L_1 \) penalty function satisfies the Goldstein-Armijo conditions, we show that the algorithm of [18] produces a point satisfying the first-order optimality conditions (4) with a fixed \( \mu \). This result also establishes the global convergence of the interior-point method algorithm using the modified search direction in the endgame.

In Section 2, we introduce the notion of modified search directions. Based on these ideas, we propose an algorithm utilizing the modified search directions in Section 3. We prove convergence of our algorithm in Section 4.

1.1. Notation and Assumptions

We introduce the following notation used throughout the article. Let \( e \) be the vector in which all elements are ones, and let \( e_i \) be the vector with all zeros except that the \( i \)th component is equal to one. Let \( K \) denote the matrix \( A\Theta A^T \). If \( C \) is a square matrix, \( \text{diag}(C) \) is the vector formed from the main diagonal of \( C \); if \( v \) is a vector, \( \text{diag}(v) \) is a diagonal matrix with the elements of \( v \) on the main diagonal.

The variables \( x_j, y_j, \) and \( z_j \) denote the \( j \)th vector in the sequence \( \{x_j\}, \{y_j\}, \) and \( \{z_j\} \), respectively. The Greek variable \( \chi_i \) denotes the \( i \)th component of the vector \( x_j \), where the index of \( x \) will be clear from the context, i.e., \( x_j = (\chi_1, \ldots, \chi_n)^T \). Similarly, we define \( y_j = (\eta_1, \ldots, \eta_m)^T \) and \( z_j = (\zeta_1, \ldots, \zeta_m)^T \) for \( y_j \in \mathbb{R}^m \) and \( z_j \in \mathbb{R}^n \).

The solution of (4) for a fixed \( \mu \) is denoted as \( x^*(\mu), y^*(\mu), \) and \( z^*(\mu) \). Capital letters \( X, Y, \) and \( Z \) denote diagonal matrices containing vectors \( x, y, \) and \( z \) on the main diagonals respectively. Let \( S_Y = \{ y \in \mathbb{R}^m \mid \|y\| \leq \Lambda_Y \} \) and \( S_Z = \{ z \in \mathbb{R}^n \mid 0 < \Omega_Z e \leq z \leq \Lambda_Z e \} \), where \( \Lambda_Y, \Omega_Z, \Lambda_Z \) are positive numbers. Furthermore, we assume that \( X_b = \{ x \in \mathbb{R}^n \mid Ax = b, x \geq 0 \} \) is compact.

2. MODIFIED SEARCH DIRECTIONS FOR THE ENGAME

In this section, we discuss the motivation for modifying search directions in the final iterations of interior-point methods.

The algorithm in [17] adaptively determines preconditioners by either computing a Cholesky factorization \( A\Theta A^T = LPL^T \) (\( L \) being an \( m \times m \) unit lower triangular matrix and \( P \) being diagonal) or by applying \( \alpha \) rank-1 updates to a previous factorization. The rank-1 update procedure is based on the observation that

\[
A\hat{\Theta}A^T = LPL^T + \sum_{i=1}^{n} \Delta \Theta_{ii} a_i a_i^T \approx LPL^T + \sum_{\alpha \text{ largest } |\Delta \Theta_{ii}|} \Delta \Theta_{ii} a_i a_i^T, \tag{9}
\]

where \( \Delta \Theta \) is the difference between the current \( \hat{\Theta} \) and the previous \( \Theta \), satisfying \( A\Theta A^T = LPL^T \), and \( a_i \) is the \( i \)th column of \( A \). If we apply \( \alpha \) rank-1 updates, we thus have a factorization of a matrix that differs from \( A\hat{\Theta}A^T \) by a matrix of rank \( (n - \alpha) \), but if \( \alpha \) is big enough to include all of the large \( |\Delta \Theta_{ii}| \) terms, the difference between \( A\hat{\Theta}A^T \) and the updated factor can be expressed as a matrix of small norm plus one of small rank. The updated factor thus can be expected to be an efficient preconditioner in the beginning and middle stage of interior-point methods. In
the course of endgame, however, a very ill-conditioned $\Theta$ may occur, and this causes difficulty in computing a good preconditioner.

Now we elaborate on the difficulty of determining good preconditioners in the later stage when iterates $x$, $y$, and $z$ approach the solution of the first-order optimality conditions. For each $i$, the strict complementarity condition forces either $x_i$ or $\zeta_i$ to go to zero [8]. Since the $i^{th}$ diagonal entry of matrix $\Theta$ is $\Theta_{ii} = x_i/\zeta_i$, the matrix $\Theta$ can contain some very small positive entries and some very large entries. These large entries correspond to small $\zeta_i$ and only these large entries are significant. On the other hand, the wildly changing entries may hamper the preconditioned conjugate gradient solver.

To overcome the difficulty, we may modify the search directions slightly by perturbing small entries of the dual slack variable $z$ in the left hand side of equation (5), so that the resulting perturbed system can be easily solved by preconditioned conjugate gradients. To elaborate, we first introduce some notation. Assume that we have a Cholesky factorization $A\Theta A^T = LPL^T$. We partition the diagonal entries of $\Theta$ into $[\Theta^B, \Theta^S]$, where $\Theta^B$ and $\Theta^S$ contain the big and small entries in $\Theta$, respectively. The current $\Theta$ is then partitioned compatibly as $[\hat{\Theta}^B, \hat{\Theta}^S]$. Then we slightly perturb the small $\zeta_i$, so that the perturbed matrix $\hat{\Theta}^B = \tau\Theta^B$, where $\tau$ is a constant. That is, we suitably choose $\varepsilon_i \in \mathbb{R}$ for every $\hat{\Theta}_{ii} \in \hat{\Theta}^B$ to obtain positive perturbed entries

\[
\hat{\zeta}_i = \zeta_i + \varepsilon_i,
\]

such that

\[
\hat{\Theta}_{ii} = \frac{x_i}{\hat{\zeta}_i} = \frac{x_i}{\zeta_i + \varepsilon_i} = \tau\Theta_{ii}.
\]

After the modification, the perturbed system becomes

\[
A\hat{\Theta} A^T = \sum \hat{\Theta}_{ii}^B a_i a_i^T + \sum \hat{\Theta}_{ii}^S a_i a_i^T = \tau LPL^T + \sum \Delta \hat{\Theta}_{ii}^S a_i a_i^T,
\]

where $\Delta \hat{\Theta}_{ii}^S = \hat{\Theta}_{ii}^S - \tau\Theta_{ii}$. Using $LPL^T$ as a preconditioner of (12), we have

\[
(LPL^T)^{-1} (A\hat{\Theta} A^T) = \tau I + (LPL^T)^{-1} \sum \Delta \hat{\Theta}_{ii}^S a_i a_i^T.
\]

It is clear that only a small number of terms in the summation in (13) can be of significant size, and thus, the preconditioned matrix is the identity plus a matrix of small rank plus a matrix of small norm; consequently, the preconditioned conjugate gradient method will converge rapidly.

3. THE MAIN ALGORITHM

The discussion in Section 2 leads to the main algorithm in this section. The merit function used in the algorithm is

\[
M(x, \rho) = c^T x - \mu \sum \ln x_i + \rho \|Ax - b\|_1,
\]

where $\rho$ is a sufficiently large finite positive number. The inner loop of the algorithm solves a barrier subproblem (2) corresponding to a given $\mu$. This is done by solving a sequence of perturbed KKT systems to determine the search directions. The $j^{th}$ system in the sequence is

\[
\begin{pmatrix}
X_j & \hat{Z}_j & 0 \\
I & 0 & \hat{A}^T \\
0 & A & 0 \\
\end{pmatrix}
\begin{pmatrix}
\Delta z_j \\
\Delta x_j \\
\Delta y_j \\
\end{pmatrix}
= - \begin{pmatrix}
X_j z_j - \mu e \\
A^T y_j + z_j - c \\
A x_j - b \\
\end{pmatrix}.
\]
The diagonal matrix $\tilde{Z}_j$ is created by perturbing diagonal entries in $Z$ that are smaller than a user-defined positive threshold $\nu_j$. We also require that the perturbed entries $\bar{\xi}_i$ be in the set $S_Z$. Our scheme for constructing $\tilde{Z}_j$ is described in detail in the first box of the inner loop in Algorithm 1. We also modify the $y_j$ so that $\eta_i \in S_Y$. Other parts of the algorithm are similar to standard interior-point methods.

Algorithm 1 is similar to an algorithm proposed by Gill, Murray, Poncelet and Saunders [19]. Their algorithm allows the variables $y$ and $z$ to be chosen arbitrarily within two bounded sets. Our algorithm, in contrast, explicitly states the way we choose $y$ and $z$ and allows the matrix $\tilde{Z}$ in the Jacobian matrix of (4) to differ from the vector $z$ in the gradient of (4).

Although Algorithm 1 and the one presented in [18] are both motivated from the discussion in Section 2, slight differences exist between the two algorithms. For each $\mu$, Algorithm 1 suitably perturbs small $z$ and solves the perturbed KKT system exactly (in the asymptotic sense); while the algorithm in [18] modifies large entries in $\Theta$ and then solves the resulting normal equations approximately. Furthermore, while practical implementation details and promising numerical results are presented in [18], convergence analysis was absent. We thus focus on the convergence analysis of the algorithm in the following sections.

**Algorithm 1. The IPM with Modified Search Directions.**

<table>
<thead>
<tr>
<th>% initialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialize $k \leftarrow 0$, $\mu_0 &gt; 0$, $\Omega_Z &gt; 0$, $\Lambda_Z &gt; 0$, $\Lambda_Y &gt; 0$, $x(\mu_0) &gt; 0$, $y(\mu_0) \in S_Y$, $z(\mu_0) \in \tilde{S}_Z$.</td>
</tr>
</tbody>
</table>

% outer loop
until (the relative duality gap is small) do

% inner loop
until (converge to the optimality conditions (4)) do

<table>
<thead>
<tr>
<th>Determine the perturbed vector $\tilde{z}_j$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose $\nu_j$ satisfying $\Omega_Z &lt; \nu_j &lt; 1$ and $\tau_j &gt; 0$.</td>
</tr>
<tr>
<td>Determine $\tilde{z}_j = (\tilde{\xi}_1, \cdots, \tilde{\xi}_n)^T$ such that</td>
</tr>
<tr>
<td>$\tilde{\xi}<em>i = \tau_j \times$ (previous $\Theta</em>{ii}$ involving refactor), if $\xi_i \leq \nu_j$,</td>
</tr>
<tr>
<td>$\tilde{\xi}_i = \xi_i$, otherwise.</td>
</tr>
<tr>
<td>Modify $\tilde{z}_j$, if necessary, so that</td>
</tr>
<tr>
<td>$\tilde{\xi}_i = \Omega_Z$, if $\tilde{\xi}_i &lt; \Omega_Z$,</td>
</tr>
<tr>
<td>$\tilde{\xi}_i = \Lambda_Z$, if $\Lambda_Z &lt; \tilde{\xi}_i$.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Determine the vector $y_j \in S_Y$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose $y_j$ such that</td>
</tr>
<tr>
<td>$\eta_i = \eta_i$, if $</td>
</tr>
<tr>
<td>$\eta_i = \frac{\Lambda_Y}{m}$, otherwise.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Update the vectors $x_j$, $y_j$, and $z_j$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determine $\Delta x_j$, $\Delta y_j$, and $\Delta z_j$ by solving equation (15).</td>
</tr>
<tr>
<td>Determine $\alpha_j$ and $x_{j+1} = x_j + \alpha_j \Delta x_j$ such that</td>
</tr>
<tr>
<td>$M(x_{j+1}, \rho) &lt; M(x_j, \rho)$.</td>
</tr>
<tr>
<td>Compute $y_{j+1} = y_j + \Delta y_j$ and $z_{j+1} = z_j + \Delta z_j$.</td>
</tr>
<tr>
<td>Set $j \leftarrow j + 1$.</td>
</tr>
</tbody>
</table>

end until

Set $x(\mu_{k+1}) = x_j$, $y(\mu_{k+1}) = y_j$, $z(\mu_{k+1}) = z_j$, and $k \leftarrow k + 1$.

Choose a positive $\mu_k < \mu_{k-1}$.

end until
4. CONVERGENCE ANALYSIS

We now prove the global convergence of Algorithm 1. We will focus on proving that an iterative method converges to the solution of (4) for a fixed $\mu_k$ in the inner loop of Algorithm 1. The global convergence of the interior-point method with modified search direction (Algorithm 1) then follows directly from the classical results by Fiacco and McCormick [20]. Note that we adopt the procedure described by Gill et al. [19] to prove the convergence of Algorithm 1.

We first state two lemmas to be used for proving the main convergence theorem. Lemma 2 shows that a certain level set is compact and Lemma 3 states that if $x \in S$, then $x$ is uniformly bounded away from zero due to the properties of $B(x, \mu)$.

**Lemma 2.** Let $B(x, \mu) = c^T x - \mu \sum_{i=1}^n \ln x_i$ and $M(x, \rho) = B(x, \mu) + \rho \|Ax - b\|_1$, where $\rho > 0$. For any given positive numbers $\Lambda_r$ and $\Lambda_M$, the level set $S = \{x \in \mathbb{R}^n \mid \|Ax-b\|_1 \leq \Lambda_r, \ M(x, \rho) \leq \Lambda_M\}$ is compact.

**Lemma 3.** There exist uniform lower and upper bounds $\Omega_s, \Lambda_s > 0$, such that $\Omega_s \leq x \leq \Lambda_s$ for any $x \in S$.

In Lemma 4, we show that a descent direction for the merit function $M(x, \rho)$ can be determined by solving equation (15). Note that the diagonal matrix $\tilde{Z}$ in equation (15) may have entries different from those in the vector $z$ in the right-hand side.

**Lemma 4.** Let $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\text{diag}(\tilde{z}) \in \mathbb{R}^n$, $r = (Ax - b) \in \mathbb{R}^m$, and $\Omega_Y, \Lambda_Y, \Lambda_z, \Lambda_r > 0$. Assume further that $x \in S$, $\|y\| < \Lambda_Y$, $\|z\| < \Lambda_Z$, and $\|r\| = \|Ax - b\|_1 < \Lambda_r$. If $\Delta x \in \mathbb{R}^n$ is the solution of equation (15) and $\rho$ is large enough, $\Delta x$ is a descent direction for $M(x, \rho)$ whenever either $N^T(c - \mu x^{-1} e)$ or $r = Ax - b$ is nonzero, where the columns of $N$ form a basis for the null space of $A$. Furthermore, $\Delta x$ is a descent direction for $\|Ax - b\|_1$, whenever $r = Ax - b$ is nonzero.

**Proof.** Our goal is to show that the inner product of $\Delta z$ and $\nabla_x M(x, \rho)$ is less than zero. We first confirm that $\Delta z$ and $\nabla_x M(x, \rho)$ are well defined. By eliminating $Ax$ and $\Delta y$ from equation (15), we obtain the reduced $2 \times 2$ KKT system

$$\begin{pmatrix} \tilde{Z} X^{-1} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = -\begin{pmatrix} c - \mu x^{-1} e - A^T y \\ A x - b \end{pmatrix}$$

and

$$\tilde{Z} X^{-1} \Delta x - A^T \Delta y = -g + A^T y,$$

where $g = c - \mu X^{-1} e$. The assumption $x \in S$ together with Lemmas 2 and 3 then imply that the solution $\Delta z$ is bounded. Furthermore, $\nabla_x M(x, \rho)$ is well defined for all $x > 0$ since

$$\nabla_x M(x, \rho) = \nabla_x B(x, \mu) \mid \rho A^T \bar{e} = g \mid \rho A^T \bar{e},$$

where $B(x, \mu)$ is defined in Lemma 2 and the $i$th component of $\bar{e}$ is equal to 1 or $-1$ depending on whether the $i$th component of $r$ is nonnegative or not.

We first establish three equations, (18), (19), and (20), that will be used for computing the inner product. There exists $\Delta x_N \in \mathbb{R}^{n-m}$ and $\Delta x_A \in \mathbb{R}^m$ such that

$$\Delta x = N \Delta x_N + A^T \Delta x_A.$$  

Multiplying equation (17) by $N^T$ and using the fact that $AN = 0$, we obtain

$$N^T \tilde{Z} X^{-1} \Delta x = -N^T g.$$

Decomposition (18) suggests that

$$N^T \tilde{Z} X^{-1} N \Delta x_N = -N^T \tilde{Z} X^{-1} A^T \Delta x_A - N^T g,$$
or
\[ H_N \Delta x_N = -N^T (g + \tilde{Z}X^{-1}A^T \Delta x_A), \]  
(19)
where \( H_N = N^T \tilde{Z}X^{-1}N \) is a positive definite matrix with full rank. Furthermore, equations (16) and (18) imply that
\[-r = -(Ax - b) = AA^T \Delta x_A,
\]
or
\[ \Delta x_A = -(AA^T)^{-1} r. \]
(20)
Now, by using equations (18)-(20), and the fact that \( AN = 0 \), we manipulate the product of \( \Delta x \) and \( \nabla_x M(x, \rho) \) as follows:
\[
(\Delta x)^T \nabla_x M(x, \rho) = (\Delta x_N)^T N^T (g + \rho A^T \bar{e}) + (\Delta x_A)^T A (g + \rho A^T \bar{e})
\]
\[= [-H_N^{-1}N^T (g + \tilde{Z}X^{-1}A^T \Delta x_A)]^T N^T (g + \rho A^T \bar{e}) + (\Delta x_A)^T A (g + \rho A^T \bar{e})
\]
\[= -g^T NH_N^{-1}N^T g - (\Delta x_A)^T A \tilde{Z}X^{-1}NH_N^{-1}N^T g - g^T NH_N^{-1}N^T \rho A^T \bar{e}
\]
\[+ (\Delta x_A)^T A \tilde{Z}X^{-1}N \tilde{H}^{-1}NH_N^{-1}N^T g - g^T NH_N^{-1}N^T \rho A^T \bar{e}
\]
\[= -g^T NH_N^{-1}N^T g + (\Delta x_A)^T A \tilde{Z}X^{-1}NH_N^{-1}N^T g - (\Delta x_A)^T A \tilde{Z}X^{-1}NH_N^{-1}N^T g
\]
\[= -g^T NH_N^{-1}N^T g + r^T (AA^T)^{-1} A \tilde{Z}X^{-1}NH_N^{-1}N^T g - r^T (AA^T)^{-1} Ag
\]
\[= -g^T NH_N^{-1}N^T g + r^T u - \rho r^T \bar{e},
\]
(21)
where \( u = (AA^T)^{-1}A(I - X^{-1}\tilde{Z}NH_N^{-1}N^T)g \).

Now we give upper bounds for the last three terms in equation (21). For the first term, we use the fact that, if \( H_N^{-1} \) is symmetric, the magnitude of \( g^T NH_N^{-1}N^T g \) is bounded by the largest and smallest eigenvalue of \( H_N^{-1} \) times \( \|N^T g\|_2^2 \), see [21]. We therefore obtain
\[-g^T NH_N^{-1}N^T g \leq -c_1 \|N^T g\|_2^2,
\]
where \( c_1 = 1/(\lambda_{\max}) \), \( \lambda_{\max} \) is the largest eigenvalue of \( H_N \), and \( \lambda_{\max} > 0 \) since \( H_N \) is positive definite. Moreover, to bound the summation of the second and third terms, let
\[ h(x) = r^T u = (Ax - b)^T (AA^T)^{-1} A (I - X^{-1}\tilde{Z}NH_N^{-1}N^T) g
\]
for any \( x \in S \) and a fixed \( \tilde{Z} \). Since \( h(x) \) is continuous and \( S \) is compact, there exists \( x_{\min} \in S \) minimizing \( h(x) \) over \( x \in S \). That is, \( h(x_{\min}) = \min\{h(x) \mid x \in S\} \). If \( h(x_{\min}) \geq r^T \bar{e} = \|r\|_1 \geq 0 \), then
\[-r^T u - \rho r^T \bar{e} \leq -h(x_{\min}) - \rho r^T \bar{e} \leq -\rho \|r\|_1.
\]
Otherwise, if \( h(x_{\min}) < r^T \bar{e} \), then for any \( \rho > 1 - h(x_{\min})/r^T \bar{e} \geq 0 \),
\[-r^T u - \rho r^T \bar{e} \leq -h(x_{\min}) - \rho r^T \bar{e} \leq -\|r\|_1.
\]
Thus, for a given \( \rho \geq \max(1 - r^T \bar{u}_{\min}/r^T \bar{e}, 1) \), we may choose \( c_2 = \min\{\rho, 0\} \), such that
\[-r^T u - \rho r^T \bar{e} \leq -c_2 \|r\|_1.
\]
Therefore, we conclude that
\[
(\Delta x)^T \nabla_x M(x, \rho) = -g^T NH_N^{-1}N^T g - r^T u - \rho r^T \bar{e}
\]
\[\leq -c_1 \|N^T g\|_2^2 - c_2 \|r\|_1 \leq 0,
\]
where \( c_1, c_2 > 0 \).
This completes the proof that the product \((\Delta x)^\top \nabla_x M(x, \rho)\) is strictly less than zero whenever either \(N^\top \rho\) or \(r\) is nonzero.

To see that \(\Delta x\) is a descent direction for \(\|Ax - b\|_1\), we simply compute the dot product of \(\Delta x\) and \(\nabla_x(\|Ax - b\|_1)\) by using equations (18) and (20)

\[
(\Delta x)^\top \nabla_x(\|Ax - b\|_1) = ((\Delta x)^\top N^\top + (\Delta x)^\top A) (A^\top \bar{e})
\]

\[
= (\Delta x)^\top (A A^\top) \bar{e}
\]

\[
= -r^\top \bar{e}
\]

\[
= -\|r\|_1.
\]

It is thus straightforward that if \(r \neq 0\), then \((\Delta x)^\top \nabla_x(\|Ax - b\|)\) is negative.

**Lemma 5.** If \(\Delta x\) is defined as in Lemma 4, then there exist positive numbers \(\alpha\) and \(\Lambda_x\) such that the Goldstein-Armijo sufficiency conditions [22] are satisfied with \(x + \alpha \Delta x > \Lambda S e\).

**Proof.** For simplicity, we adopt the notation \(M(x)\) instead of \(M(x, \rho)\) in this proof. Let \(\alpha_M\) be the largest feasible step length along \(\Delta x\); that is, \(x + \alpha_M \Delta x \geq 0\) and some element of \(x + \alpha_M \Delta x\) equals zero. By the continuity of \(M(x)\) and the fact that \(\Delta x\) is a descent direction for \(M(x)\) (Lemma 4), for sufficiently small \(\alpha > 0\), we have

\[
M(x + \alpha \Delta x) < M(x) + \xi \alpha \Delta x^\top \nabla_x M(x),
\]

where \(0 < \xi < 1\). Note that \(M(x + \alpha \Delta x) \to \infty\) as \(\alpha \to \alpha_M\), and \(M(x) + \xi \alpha \Delta x^\top \nabla_x M(x)\) decreases as \(\alpha\) increases, and hence, there exists \(\hat{\alpha} < \alpha_M\) such that

\[
M(x + \hat{\alpha} \Delta x) = M(x) + \xi \hat{\alpha} \Delta x^\top \nabla_x M(x).
\]

That is, the inequality

\[
M(x + \alpha \Delta x) - M(x) \leq \xi \alpha \Delta x^\top \nabla_x M(x)
\]

is true for every \(\alpha \in (0, \hat{\alpha})\).

Moreover, \(x + \alpha \Delta x > \Lambda S e\) holds by Lemma 4 and definition of the set \(S\).

**Theorem 6.** Main Theorem. Given positive constants, \(\Omega_Z\), \(\Lambda_Y\), and \(\Lambda Z\), let \(\{\bar{Z}_j\}\) be a sequence of diagonal matrices with \(0 < \Omega Z e \leq \text{diag}(\bar{Z}_j) \leq \Lambda Z e\), and let \(\{y_j\}\) be a sequence of vectors satisfying \(\|y_j\| \leq \Lambda Y\). Assume \(\{x_j\}\) is a sequence generated by \(x_{j+1} = x_j + \alpha_j \Delta x_j\) and \(x_0 > 0\), where \(\Delta x_j\) is defined by (15) and \(\alpha_j\) satisfies the Goldstein-Armijo sufficiency conditions on \(M(x_j, \rho)\) with \(x_j > 0\). If \(\rho\) is large enough, \(\lim_{j \to \infty} x_j = x^*(\mu)\).

**Proof.** We first choose \(x_0 \in \mathbb{R}\) and set \(\Lambda_z = \|Ax_0 - b\|\) and \(\Lambda M = M(x_0, \rho)\). Then \(S = \{x \mid \|Ax - b\| \leq \Lambda z, M(x, \rho) \leq \Lambda M\}\). Lemma 5 shows a step may be taken to decrease \(M(x_j, \mu)\) and \(\|Ax_j - b\|\). The sequence \(\{x_j\}\) generated by the iteration \(x_{j+1} = x_j + \alpha_j \Delta x_j\) thus belongs to \(S\). Since \(M(x, \rho)\) is continuous and \(S\) is compact, \(M(x, \rho)\) is bounded below over the set \(\{x \mid x \in S\}\). Lemma 4 also shows that \(M(x_j, \rho)\) is monotonically decreasing. The fact that \(\{M(x_j, \rho)\}\) is bounded below and monotonically decreasing implies that the sequence \(\{M(x_j, \rho)\}\) converges. That is, \(\lim_{j \to \infty} (M(x_{j+1}, \rho) - M(x_j, \rho)) = 0\).

On the other hand, Lemma 4 and equations (25) and (22) give

\[
-(M(x_j + \alpha_j \Delta x_j) - M(x_j)) \geq -\xi \alpha_j \Delta x_j^\top \nabla x M(x_j)
\]

\[
\geq \delta_1 \|N^\top g(x_j)\|^2 + \delta_2 \|r(x_j)\|_1
\]

\[
\geq 0.
\]

Therefore,

\[
\lim_{j \to \infty} (\alpha_j \Delta x_j^\top \nabla x M(x_j, \rho)) = \lim_{j \to \infty} \left(\delta_1 \|N^\top g(x_j)\|^2 + \delta_2 \|r(x_j)\|_1\right) = 0,
\]
The fact that (2) is a strictly convex problem for a given \( p \) suggests that the solution of (4), \( x^*(\mu) \), is unique. Finally by the continuity of \( N^T g(x) \) and \( r(x) \), we conclude that \( x_j \rightarrow x^*(\mu) \).

In Corollary 7, we show that \( \{y_j\} \rightarrow y^*(\mu) \) and \( \{z_j\} \rightarrow z^*(\mu) \) by applying Theorem 6.

**COROLLARY 7.** Under the assumptions of Theorem 6, \( \lim_{j \to \infty} (y_j + \Delta y_j) = y^*(\mu) \) and \( \lim_{j \to \infty} (z_j + \Delta z_j) = z^*(\mu) \), where \( y^*(\mu) \) and \( z^*(\mu) \) are the solution of equation (4).

**PROOF.** We first show that \( \lim_{j \to \infty} (z_j + \Delta z_j) = z^*(\mu) \). Equation (15) implies

\[
X_j \Delta z_j + \bar{Z}_j \Delta x_j = -X_j z_j + \mu \bar{e},
\]
or

\[
z_j + \Delta z_j = -X_j^{-1} \bar{Z}_j \Delta x_j + \mu X_j^{-1} \bar{e}.
\]

By equation (4), the fact that \( x_j \) is bounded above zero, and \( \{x_j\} \rightarrow x^*(\mu) \) (or \( \{\Delta x_j\} \rightarrow 0 \), we obtain

\[
\lim_{j \to \infty} (z_j + \Delta z_j) = \mu (X^*(\mu))^{-1} \bar{e} = z^*(\mu).
\]

To show the convergence of \( y_j \), we expand equation (15) to obtain

\[
A^T (y_j + \Delta y_j) = c - (z_j + \Delta z_j).
\]

Equation (26) implies that \( c - (z_j + \Delta z_j) \) converges to \( c - z^*(\mu) \), which is equal to \( A^T y^*(\mu) \) by equation (4). In other word, we have

\[
\lim_{j \to \infty} A^T (y_j + \Delta y_j) = A^T y^*(\mu).
\]

Finally, the assumption that \( A \) has full row rank, i.e., \( A^T \) has full column rank, implies \( \lim_{j \to \infty} (y_j + \Delta y_j) = y^*(\mu) \).

These results on the convergence behavior of the inner loop iteration in Algorithm 1 allow us to conclude that Algorithm 1 converges globally, due to classical results by Fiacco and McCormick [20].

**REFERENCES**