An Interior Newton Method for Quadratic Programming*

Thomas F. Coleman**
Jianguo Liu***

TR 93-1388
October 1993

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

*Research partially supported by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under grant DE-FG02-86ER25013.A000, and by NSF, AFOSR, and ONR through grant DMS-8920550, and by the Advanced Computing Research Institute, a unit of the Cornell Theory Center which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute.
**Computer Science Department and Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA.
***Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA.

AN INTERIOR NEWTON METHOD FOR QUADRATIC PROGRAMMING *

THOMAS F. COLEMAN† AND JIANGUO LIU‡

Abstract. Quadratic programming represents an extremely important class of optimization problem. In this paper we propose a new (interior) approach for the general quadratic programming problem. We establish that our new method is globally and quadratically convergent – published alternative interior approaches do not share such strong convergence properties for the nonconvex case. We also report on the results of preliminary numerical experiments: the results indicate that the proposed method has considerable practical potential.

* Research partially supported by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under grant DE-FG02-86ER25013.A000, and by NSF, AFOSR, and ONR through grant DMS-8920550, and by the Advanced Computing Research Institute, a unit of the Cornell Theory Center which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute.

† Computer Science Department and Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA.

‡ Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA.

1. Introduction. In this paper, we consider the following quadratic programming problem:

$$
(1.1) \qquad \begin{array}{ll} \min_{x \in \mathcal{R}^n} & \{q(x) = \frac{1}{2}x^T H x + c^T x\} \\ \text{subject to} & Ax = b \\ & l \leq x \leq u, \end{array}
$$

where $H \in \mathcal{R}^{n \times n}$ is symmetric and, in general, indefinite; $c \in \mathcal{R}^n$, $A \in \mathcal{R}^{m \times n}$, $b \in \mathcal{R}^m$, $l \in \{\mathcal{R} \cup \{-\infty\}\}^n$, $u \in \{\mathcal{R} \cup \{\infty\}\}^n$, and, without loss of generality, $l < u$. When H is indefinite, we are interested in locating a local minimizer and call such a minimizer a solution to (1.1).

Problem (1.1) has been studied intensively due to its importance in optimization: many real world problems are posed in the form (1.1). In addition, many algorithms for general nonlinear programming require successively solving subproblems of this form (e.g., "SQP" methods).

Surprisingly, problem (1.1) is very difficult to solve from a complexity point of view. Obtaining a global minimizer is NP-hard; moreover, unless sufficiently strong nondegeneracy assumptions are in force, Murty and Kabadi [16] establish that even obtaining a local minimizer is NP-hard. Nevertheless, the efficient determination of a local minimizer to (1.1), or at least a point satisfying second-order necessary conditions, is often possible in practise. Computational complexity questions regarding a bound on the number of steps, polynomial time guarantees, etc., are questions which we do not address in this paper. Vavasis [21], for example, discusses computational complexity as applied to quadratic programming and other optimization problems.

There are basically two kinds of existing approaches for solving (1.1) in the general case. The most popular strategies follow an active-set or gradient projection philosophy: a piecewise linear path is generated, following faces of the polytope defined by (1.1), e.g., [1], [2], [3], [6], [9], [10], [11], [15], [22]). These methods usually generate a finite sequence of intermediate "approximations". An alternative philosophy is to generate an infinite sequence of strictly feasible (or interior) points, converging in the limit to a local solution. There has been a recent resurgence of interest in such methods for linear programming, and to some degree quadratic programming, especially the convex case, e.g., [12], [5], [8], [17], [23], [24]). Most recently, Ye [24] proposed an affine scaling (interior) method for the general quadratic programming problem (1.1). Ye's method generates a sequence of points converging, in the limit, to a feasible point that satisfies first-order and second-order necessary optimality conditions. The convergence rate of Ye's algorithm is believed to be linear.

In this paper, we propose an interior Newton method using the trust region idea and a new scaling strategy. Our algorithm bears some similarity to Ye's in that the computational cost of each iteration is almost the same and both methods are "interior", i.e., all the iterates stay in the strict interior of the feasible region. But the methods differ in three important ways. First, our proposed method has quadratic convergence properties. Second, in the limit, the resulting linear system of our proposed algorithm is better conditioned[1]. Third, the convergence properties of our algorithm are established under weaker assumptions. Preliminary numerical experiments indicate that the new algorithm is efficient. A related work for box-constrained

[1] See the comment in Section 6 (Concluding Remarks).

minimization problems is given in [5].

Conditions for a point $x \in \mathcal{R}^n$ to be a local minimizer of (1.1) are well-known and can be phrased as follows: Let

$$\mathcal{F} \stackrel{\text{def}}{=} \{ x \in \mathcal{R}^n \, : \, Ax = b, \; l \le x \le u \} \tag{1.2}$$

be the feasible point set. If x is a solution to (1.1), then $\exists w \in \mathcal{R}^m$ such that

$$\text{feasibility:} \quad x \in \mathcal{F}, \tag{1.3}$$

$$\text{first-order:} \quad \begin{cases} (Hx + c + A^T w)_i = 0, & \text{if } l_i < x_i < u_i \\ (Hx + c + A^T w)_i \ge 0, & \text{if } x_i = l_i \\ (Hx + c + A^T w)_i \le 0, & \text{if } x_i = u_i, \end{cases} \tag{1.4}$$

$$\text{second order:} \quad H \ge 0 \text{ in } \mathcal{N}, \tag{1.5}$$

where

$$\mathcal{N} = \mathcal{N}(x) \stackrel{\text{def}}{=} \{ s \in \mathcal{R}^n \, : \, As = 0; \, s_i = 0, \; \forall i \in \mathcal{A}(x) \}, \tag{1.6}$$

and

$$\mathcal{A}(x) \stackrel{\text{def}}{=} \{ i \, : \, x_i = l_i \text{ or } x_i = u_i \}. \tag{1.7}$$

Conditions (1.3)-(1.5) are necessary optimality conditions. Sufficiency conditions are: $\exists w \in \mathcal{R}^m$ such that

$$\text{feasibility:} \quad x \in \mathcal{F}, \tag{1.8}$$

$$\text{first-order:} \quad \begin{cases} (Hx + c + A^T w)_i = 0, & \text{if } l_i < x_i < u_i \\ (Hx + c + A^T w)_i > 0, & \text{if } x_i = l_i \\ (Hx + c + A^T w)_i < 0, & \text{if } x_i = u_i, \end{cases} \tag{1.9}$$

$$\text{second order:} \quad H > 0 \text{ in } \mathcal{N}. \tag{1.10}$$

The multiplier w mentioned above is referred to as the Lagrange multiplier.

For a given pair $l \le x \le u$, $w \in \mathcal{R}^m$, let

$$g = g(x, w) = Hx + c + A^T w, \tag{1.11}$$

and define a vectors $\tilde{v} \in \mathcal{R}^n$ as follows: for $i = 1, \ldots, n$,

$$\tilde{v}(x)_i = \begin{cases} \min(x_i - l_i, u_i - x_i), & \text{if } l_i > -\infty \text{ or } u_i < \infty \\ 1, & \text{otherwise.} \end{cases} \tag{1.12}$$

We also define a map $\mathcal{M}_x = \mathcal{M}(x) : \mathcal{R}^n \longrightarrow \mathcal{R}^n$ as follows: if $\eta = \mathcal{M}_x(\xi)$, then

$$\eta_i = \begin{cases} x_i - u_i, & \text{if } \xi_i < 0 \text{ and } u_i < \infty \\ -\max(1, \tilde{v}_i), & \text{if } \xi_i < 0 \text{ and } u_i = \infty \\ x_i - l_i, & \text{if } \xi_i \ge 0 \text{ and } l_i > -\infty \\ \max(1, \tilde{v}_i), & \text{if } \xi_i \ge 0 \text{ and } l_i = -\infty. \end{cases} \tag{1.13}$$

Let

$$v \stackrel{\text{def}}{=} \mathcal{M}_x(g). \tag{1.14}$$

Clearly,

$$0 \le \tilde{v}_i \le |v_i|, \ \forall i. \tag{1.15}$$

Let ". *" denote componentwise multiplication. It is easy to see that (1.3) and (1.4) are equivalent to the nonlinear system

$$F(x,w) \stackrel{\text{def}}{=} \begin{bmatrix} v .* g \\ Ax - b \end{bmatrix} = 0, \tag{1.16}$$

which implies

$$\tilde{F}(x,w) \stackrel{\text{def}}{=} \begin{bmatrix} \tilde{v} .* g \\ Ax - b \end{bmatrix} = 0. \tag{1.17}$$

Suppose $x \in int(\mathcal{F}) \stackrel{\text{def}}{=} \{x \in \mathcal{F} : l < x < u\}$, $w \in \mathcal{R}^m$ are given. It is reasonable to look for a direction (s, s'), with respect to the variable pair (x, w), with the following three properties:

(1.18) (i) s is a feasible direction with respect to x, i.e., $A\,s = 0$;
(1.19) (ii) s is a descent direction for $q(x)$;
(1.20) (iii) (s, s') is a Newton direction with respect to (1.16) or (1.17) in a neighborhood of a solution.

In the following sections, we propose an algorithm which generates directions satisfying (1.18) - (1.20). We establish global and quadratic convergence properties.

In Section 2, we motivate the basic generic method, algorithm **Interior-Newton**. In Sections 3 and 4, we discuss the strong convergence properties of a particular implementation of the basic method, i.e., algorithm **Interior-Newton$_1$**. In Section 5, we give an accelerated version of the basic method, algorithm **Interior-Newton$_2$**, and discuss results of preliminary numerical experiments. (In the Appendix we show that the accelerated version of the basic algorithm maintains the theoretical convergence properties of the basic method.) Finally, in Section 6, we have some concluding remarks and observations.

2. The Basic Algorithm. The basic algorithm can be motivated in the following way. Let $\hat{v}$ be either v or $\tilde{v}$ and $\hat{F}$ be either F or $\tilde{F}$. Suppose $x \in int(\mathcal{F})$, $w \in \mathcal{R}^m$ are given such that at (x, w), $\hat{F}$ is differentiable and $\nabla \hat{F}$ is nonsingular. The Newton direction with respect to $\hat{F} = 0$ is defined as the solution to

$$\nabla \hat{F}\, p' = -\hat{F}, \tag{2.1}$$

where (Matlab notation $diag(\cdot)$ will be used to denote a diagonal matrix whose diagonal elements are the specified vector)

$$\nabla \hat{F} = \begin{bmatrix} diag(g) * (\nabla_x \hat{v}) + diag(\hat{v})H & diag(\hat{v})A^T \\ A & 0 \end{bmatrix}, \quad p' = \begin{bmatrix} p \\ p_w \end{bmatrix}. \tag{2.2}$$

System (2.2) leads to the approximation

$$\text{(2.3)} \qquad \begin{cases} Mp + A^T p_w &= -g \\ Ap &= 0, \end{cases}$$

where

$$\text{(2.4)} \qquad M \;=\; H + (diag(\hat{v})^{-1} * (\nabla_x \hat{v}) * diag(sign(g))) * diag(|g|).$$

Motivated by (2.4) and in the interests of generality, let us assume that M is a matrix of the form

$$\text{(2.5)} \qquad M = H + D^{-2}\, diag(|g|).$$

Clearly, in some neighborhood of a solution where the sufficiency conditions hold, $diag(\hat{v})^{-1} * (\nabla_x \hat{v}) * diag(sign(g)) \geq 0$, and (2.4) is recovered from (2.5) if $D = (diag(\hat{v})^{-1} * (\nabla_x \hat{v}) * diag(sign(g)))^{\frac{1}{2}}$; however, for the remainder of this section we assume only that M has the form (2.5) and D is a positive diagonal matrix.

Assume Z is a matrix such that $null(A) =< Z >$, where $< \cdot >$ denotes the subspace spanned by the columns of a matrix. When $Z^T M Z$ is nonsingular and A is of full row rank, we may solve (2.3) as follows: first, we solve

$$\text{(2.6)} \qquad Z^T M Z y = -Z^T g$$

for y, then

$$\text{(2.7)} \qquad p = Zy, \quad \text{and} \quad p_w = -(AA^T)^{-1} A(Mp + g).$$

To ensure a well-defined descent direction, we solve

$$\text{(2.8)} \qquad \min_y \{ y^T Z^T g + \frac{1}{2} y^T Z^T M Z y \; : \; \|D^{-1} Z y\| \leq \Delta \},$$

and let $p = Zy$, where D is a positive diagonal scaling matrix and $\Delta \in [\Delta_l, \Delta_u]$ for some given $0 < \Delta_l < \Delta_u$. Equivalently, the direction p is the solution to

$$\text{(2.9)} \qquad \min_p \{ \psi(p) \stackrel{\text{def}}{=} p^T g + \frac{1}{2} p^T M p \; : \; Ap = 0, \|D^{-1} p\| \leq \Delta \}.$$

Note that subproblems (2.8) and (2.9) have the disturbing property that some elements of M may approach infinity as some diagonal components of D go to zero. However, substitution of the scaling matrix D can ameliorate this problem. Specifically, let $\bar{Z}$ be a matrix with orthogonal columns such that

$$\text{(2.10)} \qquad < \bar{Z} >= null(\bar{A}), \text{ where } \bar{A} = AD.$$

Assume that $D^{-1} Z$ has full column rank, we solve

$$\text{(2.11)} \qquad \min_{\bar{p}} \{ \bar{\psi}(\bar{p}) \stackrel{\text{def}}{=} \bar{p}^T \bar{Z}^T \bar{g} + \frac{1}{2} \bar{p}^T \bar{Z}^T \bar{M} \bar{Z} \bar{p} \; : \; \|\bar{p}\| \leq \Delta \},$$

where

$$\text{(2.12)} \qquad \bar{M} = DMD = DHD + diag(|g|), \text{ and } \bar{g} = Dg.$$

Then

$$p = D\bar{Z}\bar{p}. \tag{2.13}$$

Note that all the elements of $\bar{M}$ are bounded. Also,

$$\psi(p) = \bar{\psi}(\bar{p}). \tag{2.14}$$

It is well known (see, e.g., [20]) that if $\bar{p}$ is the solution to (2.11), then $\exists\lambda \geq 0$ such that

$$(\bar{Z}^T\bar{M}\bar{Z} + \lambda I)\bar{p} = -\bar{Z}^T\bar{g}, \tag{2.15}$$

$$\bar{Z}^T\bar{M}\bar{Z} + \lambda I \geq 0, \tag{2.16}$$

and

$$\lambda(\Delta - \|\bar{p}\|) = 0. \tag{2.17}$$

By the definition of $\bar{Z}$, $\exists p_w \in \mathcal{R}^m$ such that

$$(\bar{M} + \lambda I)\bar{Z}\bar{p} + \bar{g} + \bar{A}^T p_w = 0, \tag{2.18}$$

or equivalently,

$$(M + \lambda D^{-2})p + g + A^T p_w = 0. \tag{2.19}$$

If $(\bar{A}\bar{A}^T)^{-1}$ exists, then

$$p_w = -(\bar{A}\bar{A}^T)^{-1}\bar{A}\,(\,(\bar{M} + \lambda I)\bar{Z}\bar{p} + \bar{g}\,), \tag{2.20}$$

which is the least square solution to (2.18).

When moving in the direction p, starting from a feasible point, a variable may reach a bound. Therefore, to maintain strict feasibility and yet allow the solution (which may have some of the variables at their bounds) to be approached asymptotically (sufficiently fast) we define a "step back" procedure as follows. For each k, compute β^k and α^k by a function *stepback*:

$$[\beta^k,\, \alpha^k] = stepback(x^k, p^k, \theta^k, l, u, \tau_1, \tau_2), \tag{2.21}$$

then,

$$x^{k+1} = x^k + \alpha^k p^k, \tag{2.22}$$

where[2]

$$\theta^k \stackrel{\text{def}}{=} \frac{\|v^k.*g^k\| + |\psi^k(p^k)|}{1 + \|v^k.*g^k\| + |\psi^k(p^k)|}, \tag{2.23}$$

and $\tau_1 \in (0,\, 1)$ is given, and $\tau_2 = 1$.

[2] By Lemma 3.10, $\theta = 0$ if and only if the first-order and second-order necessary optimality conditions are satisfied. So θ is a good measure of the optimality.

The function *stepback* is defined as follows:

function stepback

$function\ [\beta,\, \alpha] = stepback\,(x, p, \theta, l, u, \tau_1, \tau_2, k)$

$$
\begin{aligned}
&(2.24) \qquad & BR &= \max\{(l-x)./p,\ (u-x)./p\}, \\
&(2.25) \qquad & \beta &= \min_{1\le i\le n}\{BR_i\}, \\
&(2.26) \qquad & \rho &= \max(\tau_1, 1-\theta), \quad (if\ k=0\ ,\ set\ \rho=\tau_1) \\
&(2.27) \qquad & \alpha &= \min\,(\tau_2,\ \rho\beta).
\end{aligned}
$$

A possible way to update the multiplier is based on the fact that at a local minimizer,

$$diag(\tilde{v})\,(Hx + c + A^T w) = \tilde{v}.*g = 0. \tag{2.28}$$

So if $A \cdot diag(\tilde{v}) \cdot A^T$ is nonsingular, the multiplier can be expressed as

$$w = -(Adiag(\tilde{v})A^T)^{-1}Adiag(\tilde{v})(Hx + c). \tag{2.29}$$

Hence we may compute

$$w^k = -(A \cdot diag(\tilde{v}^k) \cdot A^T)^{-1} A \cdot diag(\tilde{v}^k)(Hx^k + c), \ \ \forall k. \tag{2.30}$$

In this case,

$$
\begin{aligned}
g &= Hx + c + A^T w \\
&= (I - A^T(A \cdot diag(\tilde{v}) \cdot A^T)^{-1} A \cdot diag(\tilde{v}))\,(Hx + c)
\end{aligned} \tag{2.31}
$$

which is a projection of $\nabla q(x) = Hx + c$ into the null space of $A \cdot diag(\tilde{v})$, and $diag(\tilde{v}^{\frac{1}{2}}) \cdot g$ is the orthorgonal projection of $diag(\tilde{v}^{\frac{1}{2}}) \cdot (Hx + c)$ into the null space of $A \cdot diag(\tilde{v}^{\frac{1}{2}})$.

Global convergence properties require sufficient reduction in the objective function. In general, the truncated step $\alpha^k p^k$, where p^k solves (2.9) and α_k is determined by function **stepback**, does not guarantee sufficient reduction; therefore, similar to the strategy used in [5], a (truncated) scaled gradient computation is required to "augment" the truncated step $\alpha^k p^k$. To this end let w solve the following equation:

$$AD^2 g = AD^2(Hx + c + A^T w) = 0, \tag{2.32}$$

which is satisfied if AD^2A^T is nonsingular and w is the solution to

$$(AD)^T w \overset{\text{LS}}{=} -D(Hx + c). \tag{2.33}$$

Let

$$y = \frac{\bar{g}}{\|\bar{g}\|}, \tag{2.34}$$

and

$$p_g = \mu D y = \frac{\mu D \bar{g}}{\|\bar{g}\|}, \tag{2.35}$$

where, to match with the formulation in (2.11), μ is defined as a solution to the following problem:

$$\min_{\mu\in\mathcal{R}}\{\psi_g(\mu) \stackrel{\text{def}}{=} \mu y^T\bar{g} + \frac{\mu^2}{2} y^T\bar{M}y \ : \ |\mu| \le \Delta\,\}. \tag{2.36}$$

It is clear that

$$\psi(p_g) = \psi_g(\mu), \quad Ap_g = 0, \tag{2.37}$$

and, similar to (2.15) – (2.17), $\exists\lambda_g \ge 0$ such that

$$(y^T\bar{M}y + \lambda_g)\mu = -y^T\bar{g}, \tag{2.38}$$

$$y^T\bar{M}y + \lambda_g \ge 0, \tag{2.39}$$

and

$$\lambda_g(\Delta - |\mu|) = 0. \tag{2.40}$$

Also, we may compute β_g and α_g using function *stepback*:

$$[\beta_g^k,\, \alpha_g^k] = \mathit{stepback}\,(x^k, p_g^k, \theta^{k-1}, l, u, \tau_1, \tau_2), \tag{2.41}$$

where τ_1, $\tau_2 \in (0,\,1)$ are given. We use θ^{k-1} here because $\psi^k(p^k)$ may not be available yet when we compute p_g^k. Note that by (2.27),

$$\alpha_g^k \le \tau_2 < 1, \ \ \forall k. \tag{2.42}$$

We can now state the basic generic algorithm.

Algorithm Interior-Newton

Let $x^0 \in int(\mathcal{F})$ be given.

For $k = 0, 1, 2, \cdots$

1. Determine D^k,
   and compute w^k, e.g., by (2.33).
2. Compute s^k such that $x^k + s^k \in int(\mathcal{F})$,
   e.g., s^k is an affine combination of $\alpha^k p^k$ and $\alpha_g^k p_g^k$,
   defined by (2.13), (2.27), (2.35) and (2.41), respectively.
3. Update $x^{k+1} = x^k + s^k$.

Algorithm **Interior-Newton** allows for considerable freedom in its definition; for example, the positive diagonal matrix D^k is unspecified. Two natural choices are $D^k = (\tilde{V}^k)^{\frac{1}{2}}$ and $D^k = |V^k|^{\frac{1}{2}}$. We consider each choice in turn. In the next two sections we consider the convergence properties of algorithm **Interior-Newton** when $D^k = (\tilde{V}^k)^{\frac{1}{2}}$: we call this version of the basic algorithm, **Interior-Newton$_1$**.

3. Convergence of Algorithm Interior-Newton$_1$. In this section we establish global convergence properties of algorithm **Interior-Newton** when the diagonal scaling matrix, D^k, is chosen $D^k = (\tilde{V}^k)^{\frac{1}{2}}$; **Interior-Newton$_1$** denotes this specification of the basic algorithm.

In this and the following section, the notations $\{s^k\}$, $\{x^k\}$, w^k, and $\{D^k = (\tilde{V}^k)^{\frac{1}{2}}\}$ are reserved for the sequences in algorithm **Interior-Newton$_1$**. In each iteration,

$$\Delta^k \in [\Delta_l, \Delta_u], \tag{3.1}$$

where $0 < \Delta_l < \Delta_u < \infty$ are two given scalars.

We will drop the superscript k sometimes when there is no confusion. All the norms considered here are $\|\cdot\|_2$ except where otherwise specified. When needed, we denote $f^k = f(x^k)$ (or $f(x^k, w^k)$) for any function $f(x)$ (or $f(x, w)$). Sometimes, we use the capital letter to denote $diag(\cdot)$ for a given vector. For example, $\tilde{V} \stackrel{\text{def}}{=} diag(\tilde{v})$ and $|G| = diag(|g|)$.

The following lemma defines a few basic equalities used throughout the remainder of this paper.

LEMMA 3.1. *Let $\bar{p}$, p and β be defined by (2.11), (2.13) and (2.25). Let μ, y, p_g and β_g be defined by (2.36), (2.34), (2.35) and (2.41). Then*

$$q(x) - q(x+s) = -\psi(s) + \frac{1}{2}s^T D^{-2}|G|s; \tag{3.2}$$

$$\psi(\sigma p) = \bar{\psi}(\sigma\bar{p}) = -\sigma(1-\frac{\sigma}{2})\bar{p}^T(\bar{Z}^T\bar{M}\bar{Z} + \lambda I)\bar{p} - \frac{\sigma^2}{2}\lambda\|\bar{p}\|^2 \le 0, \quad \forall\sigma \in [0, \min(1,\beta)]; \tag{3.3}$$

$$\psi(\sigma p_g) = \psi_g(\sigma\mu) = -\sigma(1-\frac{\sigma}{2})(y^T\bar{M}y + \lambda_g)\mu^2 - \frac{\sigma^2}{2}\lambda_g\mu^2 \le 0, \quad \forall\sigma \in [0, \min(1,\beta_g)]. \tag{3.4}$$

Proof. (i) is a direct consequence of the definition of ψ. (ii) and (iii) are true by (2.15), (2.16), (2.38), and (2.39). ∎

Throughout this paper, the following two assumptions will be in force.

(AS1) The level set $\mathcal{L} = \{x \in \mathcal{F} : q(x) \le q(x^0)\}$ is compact.
(AS2) $A\tilde{V}A^T$ is nonsingular $\forall x \in \mathcal{L}$.

The assumption **(AS1)** is standard and condition **(AS2)** is a nondegeneracy assumption. By **(AS1)** and **(AS2)**, $\exists C_1 > 0$ such that

$$\|(A\tilde{V}A^T)^{-1}\| \le C_1, \quad \forall x \in \mathcal{L}, \tag{3.5}$$

By (1.15) and **(AS2)**, we see that $A|V|A^T$ is nonsingular and we may choose C_1 sufficiently large so that

$$\|(A|V|A^T)^{-1}\| \le C_1, \quad ,\forall x \in \mathcal{L}, \forall w \in \mathcal{R}^m. \tag{3.6}$$

To distinguish from the computed w^k, for any $x \in \mathcal{L}$, we let $w_{ls} = w_{ls}(x)$ denote the solution to

$$(AD)^T w \stackrel{\text{LS}}{=} -D(Hx + c), \tag{3.7}$$

and

$$g_{ls} = g_{ls}(x) = Hx + c + A^T w_{ls}. \tag{3.8}$$

For a specific x^*, we write

$$w_{ls}^* \stackrel{\text{def}}{=} w_{ls}(x^*) \text{ and } g_{ls}^* \stackrel{\text{def}}{=} Hx^* + c + A^T w_{ls}^*. \tag{3.9}$$

By **(AS2)**, we have

$$w_{ls}^k = -(A\tilde{V}^k A^T)^{-1} A\tilde{V}^k (Hx^k + c), \ \forall k. \tag{3.10}$$

Accordingly,

$$g_{ls}^k = Hx^k + c + A^T w_{ls}^k, \tag{3.11}$$

while

$$g^k = Hx^k + c + A^T w^k. \tag{3.12}$$

A condition we need for $\{w^k\}$ is as follows:

(CD1) We say that $\{w^k\}$ satisfies **(CD1)** if $w^{k_j} \longrightarrow w_{ls}^*$ whenever $x^{k_j} \longrightarrow x^* \in \mathcal{C}$,

where $\mathcal{C}$ is a complementarity set which is defined by

$$\mathcal{C} \stackrel{\text{def}}{=} \{x \in \mathcal{F} : \tilde{v} .* (Hx + c + A^T w) = 0 \text{ for some } w \in \mathcal{R}^m\}. \tag{3.13}$$

Clearly **(CD1)** can be satisfied in practise. For example, choose $w^k = w_{ls}^k$ for all k.

To ensure convergence we need to guarantee that the reduction in the quadratic model ψ^k is a fraction of the reduction obtained by the (truncated) scaled gradient step. The following condition formalizes this.

Let $\gamma > 0$ be given.

(CD2) $\left\{\begin{array}{c} \text{We say that a vector } \xi \text{ satisfies } \mathbf{(CD2)} \text{ if} \\ \psi^k(\xi) \le \gamma\, \psi^k(\alpha_g^k p_g^k), \\ \text{where } \alpha_g^k \text{ and } p_g^k \text{ are defined by (2.41) and (2.35).} \end{array}\right.$

We note that similar conditions have been used in [5] and [14] for box-constrained and unconstrained problems respectively.

THEOREM 3.2. *Let $\{s^k\}$ be generated by Algorithm* **Interior-Newton$_1$**. *Suppose s^k satisfies* **(CD2)** *for all k sufficiently large. Then $\{q(x^k)\}$ converges and*

$$D^k g_{ls}^k \longrightarrow 0. \tag{3.14}$$

Proof. For all k sufficiently large, by (3.2), **(CD2)** and (3.4),

$$q(x^k) - q(x^{k+1}) \ge -\gamma\, \psi^k(\alpha_g^k p_g^k) \ge 0. \tag{3.15}$$

So $\{q(x^k)\}$ is monotonically decreasing, therefore, $\{q(x^k)\}$ converges by **(AS1)**. To show (3.14), using (3.15) and (3.4) again, we have

$$\alpha_g^k(1-\frac{(\alpha_g^k)^2}{2})\,((\,y^k)^T\bar{M}^k y^k+\lambda_g^k)\,(\mu^k)^2 \longrightarrow 0. \tag{3.16}$$

Using (2.38) and (2.34), we have

$$(y^T\bar{M}y+\lambda_g)\mu = -y^T D g_{ls} = -\|Dg_{ls}\|. \tag{3.17}$$

So if (3.14) is false, then $\exists\epsilon>0$ and a subsequence $\{k_j\}$ such that

$$\|D^{k_j}g_{ls}^{k_j}\| \geq \epsilon. \tag{3.18}$$

Applying (3.17), we see that $((y^{k_j})^T\bar{M}^{k_j}y^{k_j}+\lambda_g^{k_j})(\mu^{k_j})^2$ is bounded away from zero. Hence (3.16) and (2.27) imply

$$\alpha_g^{k_j} \longrightarrow 0. \tag{3.19}$$

Therefore, by (2.27), $\beta_g^{k_j} \longrightarrow 0$. Since n is finite, from the definition of β_g, we may (without loss of generality) assume that $\beta_g^{k_j}=\frac{u_1-x_1^{k_j}}{(p_g^{k_j})_1}$. Note that $\|p_g^k\|\leq\Delta_u\|D^k\|$ is bounded by (2.35) and **(AS1)**, so $u_1-x_1^{k_j} \longrightarrow 0$ and when k_j is sufficiently large, $\tilde{v}_1^{k_j}=u_1-x_1^{k_j} \longrightarrow 0$. But by (2.35),

$$\frac{\|D^{k_j}g_{ls}^{k_j}\|}{\mu^{k_j}(g_{ls}^{k_j})_1}=\frac{\tilde{v}_1^{k_j}}{(p_g^{k_j})_1} \longrightarrow 0. \tag{3.20}$$

Therefore, we must have $\|D^{k_j}g_{ls}^{k_j}\| \longrightarrow 0$ since $|\mu^{k_j}(g_{ls}^{k_j})_1|$ is bounded above. This contradicts (3.18). ∎

COROLLARY 3.3. *Let $\{x^k\}$ be generated by Algorithm* **Interior-Newton$_1$**. *Suppose x^* is any limit point of $\{x^k\}$ and $x^{k_j} \longrightarrow x^*$. If s^k satisfies* **(CD2)** *for all k sufficiently large, then*

$$\tilde{v}^*.*g_{ls}^*=\tilde{v}^*.*(Hx^*+c+A^Tw_{ls}^*)=0. \tag{3.21}$$

Furthermore, if $\{w^k\}$ satisfies **(CD1)**, *then*

$$w^{k_j} \longrightarrow w_{ls}^*. \tag{3.22}$$

Proof. It is clear that $w_{ls}^{k_j} \longrightarrow w_{ls}^*$. So by Theorem 3.2, (3.21) holds, and then, (3.22) is true by **(CD1)**. ∎

Next we show that $\{x^k\}$ is convergent. We define

$$\mathcal{S}=\{x\,:\,x \text{ is a limit point of } \{x^k\}\}, \tag{3.23}$$

and, $\forall x^*\in\mathcal{S}$,

$$\mathcal{B}^* \;=\; \mathcal{B}(x^*)=\{x\in\mathcal{F}\,:\,x_i=x_i^*,\ \forall i\in\mathcal{A}^*\}, \tag{3.24}$$

$$\mathcal{B}_I^* \;=\; \mathcal{B}_I(x^*)=\{x\in\mathcal{B}^*\,:\,\tilde{v}_i\neq 0,\ \forall i \text{ such that } \tilde{v}_i^*\neq 0\}. \tag{3.25}$$

By Corollary 3.3 and **(AS1)**, $\mathcal{C} \supseteq \mathcal{S} \neq \emptyset$ if s^k satisfies **(CD2)** for all k sufficiently large. It is easy to see that $\mathcal{B}^*$ is the face of $\mathcal{F}$ which includes x^* while $\mathcal{B}_I^*$ is the (relative) interior of $\mathcal{B}^*$. Usually, $x^* \in \mathcal{B}_I^* \subset \mathcal{B}^*$. But when x^* is a vertex, $x^* = \mathcal{B}_I^* = \mathcal{B}^*$.

We will use the following notations: given any index set $\mathcal{I}$, any matrix $A \in \mathcal{R}^{m \times n}$, and any vector $x \in \mathcal{R}^n$, let $A_{\mathcal{I}}$ denote the submatrix of A which consists of those columns of A with their column indices in $\mathcal{I}$, and let $x_{\mathcal{I}}$ denote the subvector of x which consists of those components of x with their row indices in $\mathcal{I}$. We assume the following strict complementarity condition throughout this paper:

(AS3) $\forall x \in \mathcal{C}$, if $(Hx + c + A^T w_{ls})_i = 0$, then $\tilde{v}_i \neq 0$.

Under assumption **(AS3)**, if $x^* \in \mathcal{C}$, then

$$\mathcal{A}^{*^c} = \{i : l_i < x_i^* < u_i\} = \{i : (g_{ls}^*)_i = 0\}, \tag{3.26}$$

where $\mathcal{A}$ is defined by (1.7).

The following crucial result depends on the fact that q is a quadratic function.

LEMMA 3.4. *Suppose* $x^* \in \mathcal{C} \bigcap \mathcal{L}$. *Then* $\mathcal{B}_I^* \bigcap (\mathcal{C} \bigcap \mathcal{L}) = x^*$.

Proof. The lemma is clearly true if x^* is a vertex. Now assume x^* is not a vertex. If the lemma is false, then $\exists x^\star \in \mathcal{B}_I^* \bigcap (\mathcal{C} \bigcap \mathcal{L})$ and $x^\star \neq x^*$. Let

$$x(t) = x^* + t(x^\star - x^*), \quad \forall t \geq 0. \tag{3.27}$$

Then

$$Ax(t) = b \text{ and } x(t)_{\mathcal{A}^*} = x^*_{\mathcal{A}^*}, \quad \forall t \geq 0. \tag{3.28}$$

By **(AS2)**, $A_{\mathcal{A}^{*c}}$ has full row rank. Hence the projection matrix into $null(A_{\mathcal{A}^{*c}})$

$$P^* \stackrel{\text{def}}{=} I - A^T_{\mathcal{A}^{*c}}(A_{\mathcal{A}^{*c}} A^T_{\mathcal{A}^{*c}})^{-1} A_{\mathcal{A}^{*c}} \tag{3.29}$$

is well-defined. By (3.13), $\exists w^*$ and $w^\star$ such that

$$D^*(Hx^* + c + A^T w^*) = 0, \tag{3.30}$$
$$D^\star(Hx^\star + c + A^T w^\star) = 0. \tag{3.31}$$

Since x^*, $x^\star \in \mathcal{B}_I^*$, we have $\tilde{v}^\star_{\mathcal{A}^{*c}} > 0$ and $\tilde{v}^*_{\mathcal{A}^{*c}} > 0$. Hence

$$(Hx^* + c + A^T w^*)_{\mathcal{A}^{*c}} = 0, \tag{3.32}$$
$$(Hx^\star + c + A^T w^\star)_{\mathcal{A}^{*c}} = 0. \tag{3.33}$$

Note that $(A^T w)_{\mathcal{A}^{*c}} = A^T_{\mathcal{A}^{*c}} w$, $\forall w \in \mathcal{R}^m$, we have

$$P^*(Hx^* + c)_{\mathcal{A}^{*c}} = P^*(Hx^* + c + A^T w^*)_{\mathcal{A}^{*c}} = 0, \tag{3.34}$$
$$P^*(Hx^\star + c)_{\mathcal{A}^{*c}} = P^*(Hx^\star + c + A^T w^\star)_{\mathcal{A}^{*c}} = 0, \tag{3.35}$$

and,

$$P^*(H(x^\star - x^*))_{\mathcal{A}^{*c}} = 0. \tag{3.36}$$

It follows from (3.34), (3.36) and (3.27),

$$(3.37)\quad P^*(Hx(t)+c)_{\mathcal{A}^{*c}} = P^*(Hx^*+c)_{\mathcal{A}^{*c}} + tP^*(H(x^\star - x^*))_{\mathcal{A}^{*c}} = 0, \forall t \geq 0.$$

So, $\forall t \geq 0$, $\exists w(t) \in \mathcal{R}^m$ such that

$$(3.38)\qquad (Hx(t)+c)_{\mathcal{A}^{*c}} = -A^T_{\mathcal{A}^{*c}} w(t),$$

i.e.,

$$(3.39)\qquad (Hx(t)+c+A^T w(t))_{\mathcal{A}^{*c}} = 0,\ \ \forall t \geq 0.$$

Applying (3.39) and the second equality of (3.28), we have

$$(3.40)\qquad D(x(t))(Hx(t)+c+A^T w(t)) = 0,\ \ \forall t \geq 0.$$

Note that $x(0) = x^*$, $x(1) = x^\star$, $\tilde{v}^*_{\mathcal{A}^{*c}} > 0$, and $\tilde{v}^\star_{\mathcal{A}^{*c}} > 0$, we must have

$$(3.41)\qquad \tilde{v}(x(t))_{\mathcal{A}^{*c}} > 0,\ \ \forall t \geq 0,$$

since otherwise, there would be a $t_0 > 1$ such that $x(t_0) \in \mathcal{F}$ and $\tilde{v}(x(t_0))_i = 0$ for some $i \in \mathcal{A}^{*^c}$. But then **(AS3)** would be violated by (3.39) and (3.40). Therefore

$$(3.42)\qquad l_{\mathcal{A}^{*c}} \leq x(t)_{\mathcal{A}^{*c}} \leq u_{\mathcal{A}^{*c}},\ \ \forall t \geq 0,$$

and, by (3.28), (3.42), we have

$$(3.43)\qquad x(t) \in \mathcal{F},\ \ \forall t \geq 0.$$

Note that

$$
\begin{aligned}
(3.44)\ (Hx^*+c)^T(x^\star - x^*) &= (Hx^*+c)^T_{\mathcal{A}^{*c}}(x^\star - x^*)_{\mathcal{A}^{*c}} \quad (\text{since } x^\star_{\mathcal{A}^*} = x^*_{\mathcal{A}^*}) \\
&= -(A^T w^*)^T_{\mathcal{A}^{*c}}(x^\star - x^*)_{\mathcal{A}^{*c}} \quad (\text{by } (3.32)) \\
&= -(A^T w^*)^T(x^\star - x^*) \\
&= -(w^*)^T A(x^\star - x^*) = 0, \quad (\text{since } A(x^\star - x^*) = 0)
\end{aligned}
$$

and

$$
\begin{aligned}
(3.45)\qquad & (x^\star - x^*)^T H(x^\star - x^*) \\
= \ & (x^\star - x^*)^T_{\mathcal{A}^{*c}}(H(x^\star - x^*))_{\mathcal{A}^{*c}} \\
= \ & (x^\star - x^*)^T_{\mathcal{A}^{*c}}(-A^T(w^\star - w^*))_{\mathcal{A}^{*c}} \quad (\text{by (3.32) and (3.33)}) \\
= \ & -(A(x^\star - x^*))^T(w^\star - w^*) = 0.
\end{aligned}
$$

Therefore, by (3.27), (3.44) and (3.45), we have

$$
\begin{aligned}
(3.46)\quad q(x(t)) &= q(x^*) + t(Hx^*+c)^T(x^\star - x^*) + \frac{t^2}{2}(x^\star - x^*)^T H(x^\star - x^*) \\
&= q(x^*) \leq q(x^0),\ \ \forall t \geq 0.
\end{aligned}
$$

Hence,

$$(3.47)\qquad x(t) \in \mathcal{L},\ \ \forall t \geq 0,$$

which is a contradiction to **(AS1)**. ∎

LEMMA 3.5. *Let $\{s^k\}$ be generated by Algorithm* **Interior-Newton$_1$**. *Suppose s^k satisfies* **(CD2)** *for all k sufficiently large. Then $\mathcal{S}$ is finite and $\exists x^* \in \mathcal{S}$ such that $\mathcal{B}^* \bigcap \mathcal{S} = x^*$.*

Proof. It is clear that $\mathcal{S} \subset \mathcal{C} \bigcap \mathcal{L}$. So by Lemma 3.4, the interior of any face or edge of $\mathcal{F}$ includes at most one element of $\mathcal{S}$. Since the total number of faces, edges, and vertices of $\mathcal{F}$ is finite, $\mathcal{S}$ is finite. Choose $x^* \in \mathcal{S}$ satisfying

$$|\mathcal{A}^*| = \max_{x \in \mathcal{S}} \{|\mathcal{A}(x)|\}. \tag{3.48}$$

We now show that $\mathcal{B}^* \bigcap \mathcal{S} = x^*$. In fact, if $\mathcal{A}^* = \emptyset$, then by the definitions of $\mathcal{B}^*$ and $\mathcal{B}_I^*$, $\mathcal{B}_I^* = \mathcal{B}^*$. So by Lemma 3.4, our claim follows. The claim is certainly true if x^* is a vertex. Now suppose that $\mathcal{A}^* \neq \emptyset$ and that x^* is not a vertex. If the claim is false, then $\exists x^\star \in \mathcal{B}^* \bigcap \mathcal{S},\ x^\star \neq x^*$. By Lemma 3.4, $x^\star \notin \mathcal{B}_I^* \bigcap \mathcal{S}$. Hence, $x^\star_{\mathcal{A}^*} = x^*_{\mathcal{A}^*}$, and $\exists i \in \mathcal{A}^{*^c}$ such that $\tilde{v}_i^\star = 0$. Therefore $|\mathcal{A}^\star| > |\mathcal{A}^*|$, which is a contradiction to the choice of x^*. ∎

To establish convergence of $\{x^k\}$, we need to first establish a boundedness condition on the steps s^k. Let $\gamma_0 > 0$ be given. (Recall that by (3.1), Δ_k is bounded above and below.)

$$\textbf{(CD3)} \quad \left\{ \begin{array}{l} \text{We say that a vector } \xi \text{ satisfies } \textbf{(CD3)} \text{ if} \\ \qquad\qquad \|(D^k)^{-1}\xi\| \leq \gamma_0 \Delta^k. \end{array} \right.$$

Condition **(CD3)** is satisfied, for example, if $\xi = \alpha^k p^k$ or $\xi = \alpha_g^k p_g^k$, where α^k, p^k, α_g^k and p_g^k are defined by (2.27), (2.13), (2.41) and (2.35).

THEOREM 3.6. *Let $\{s^k\}$ be generated by Algorithm* **Interior-Newton$_1$** *and suppose s^k satisfies* **(CD2)** *and* **(CD3)** *for all k sufficiently large. Then $\{x^k\}$ converges to x^* which satisfies (3.48).*

Proof. Let x^* satisfy (3.48). By Lemma 3.5, $\mathcal{B}^* \bigcap \mathcal{S} = x^*$. If $\mathcal{A}^* = \emptyset$, then $\mathcal{B}^* = \mathcal{F}$ and $\mathcal{F} \bigcap \mathcal{S} = x^*$ and the theorem is true. So we assume that $\mathcal{A}^* \neq \emptyset$.

If $\{x^k\}$ does not converge to x^*, then let

$$\mathcal{S}' \stackrel{\text{def}}{=} \mathcal{S} - \{x^*\} \neq \emptyset, \tag{3.49}$$

$$\epsilon^* \stackrel{\text{def}}{=} \min_{i \in \mathcal{A}^*} \{|(g^*_{l_s})_i|\} > 0, \tag{3.50}$$

$$C_\gamma \stackrel{\text{def}}{=} \frac{1}{3(1+\gamma_0 \Delta_u)}. \tag{3.51}$$

Since $\mathcal{S}$ is finite and $\mathcal{B}^* \bigcap \mathcal{S} = x^*$, $\exists\ 0 < \epsilon < \epsilon^*$ such that $0 < C_\gamma \epsilon < 1$ and

$$\forall x^\star \in \mathcal{S}',\ \exists i^\star = i(x^\star) \in \mathcal{A}^*, \text{ such that } |x^\star_{i^\star} - x^*_{i^\star}| > \epsilon. \tag{3.52}$$

Therefore, $\exists K_1 > 0$ such that for each $k \geq K_1$,

$$\text{either } |x_i^k - x_i^*|^{\frac{1}{2}} < C_\gamma \epsilon, \forall i \in \mathcal{A}^*, \text{ or } |x_i^k - x_i^*| > \tfrac{2}{3}\epsilon, \text{ for some } i \in \mathcal{A}^*. \tag{3.53}$$

Also, if ϵ is sufficiently small, $\exists K_2 \geq K_1$, such that $\forall k \geq K_2$,

$$\text{if } |x_i^k - x_i^*|^{\frac{1}{2}} < C_\gamma \epsilon, \forall i \in \mathcal{A}^*, \text{ then } \tilde{v}_i^k = |x_i^k - x_i^*|, \forall i \in \mathcal{A}^*, \tag{3.54}$$

and by **(CD3)**,

$$|(s^k)_i| \le \gamma_0 \Delta_u (\tilde{v}_i^k)^{\frac{1}{2}}, \ \forall i \in \mathcal{A}^*, \text{ if } |x_i^k - x_i^*|^{\frac{1}{2}} < C_\gamma \epsilon, \ \forall i \in \mathcal{A}^*. \tag{3.55}$$

Since $x^* \in \mathcal{S}$, $\exists k_0 \ge K_2$, such that

$$|x_i^{k_0} - x_i^*|^{\frac{1}{2}} < C_\gamma \epsilon, \ \forall i \in \mathcal{A}^*. \tag{3.56}$$

Then by (3.55), we have

$$\begin{aligned} |x_i^{k_0+1} - x_i^*| &\le (1 + \gamma_0 \Delta_u)|x_i^{k_0} - x_i^*|^{\frac{1}{2}} \\ &\le \frac{1}{3C_\gamma}|x_i^{k_0} - x_i^*|^{\frac{1}{2}} \le \frac{\epsilon}{3}, \ \forall i \in \mathcal{A}^*. \end{aligned} \tag{3.57}$$

So, by (3.53),

$$|x_i^{k_0+1} - x_i^*|^{\frac{1}{2}} < C_\gamma \epsilon, \ \forall i \in \mathcal{A}^*. \tag{3.58}$$

Hence, by induction, we have that $\forall k \ge k_0$,

$$|x_i^k - x_i^*| \le |x_i^k - x_i^*|^{\frac{1}{2}} < C_\gamma \epsilon \le \frac{\epsilon}{3}, \ \forall i \in \mathcal{A}^*. \tag{3.59}$$

This is a contradiction to (3.49) and (3.52). ■

We can now establish a first-order convergence result.

LEMMA 3.7. *The sequence $\{x^k\}$ generated by Algorithm* **Interior-Newton$_1$** *converges to a first- order point*[3] *x^* if for all k sufficiently large, s^k satisfies* **(CD2)**, **(CD3)** *and the following:*

$$sign((s^k)_i) = -sign((g_{l_s}^*)_i), \ \forall i \in \mathcal{A}^*. \tag{3.60}$$

Proof. Let x^* be the point in Theorem 3.6. We only need to show that x^* satisfies (1.4). In fact, by (3.21), $\tilde{v}_i^*(g_{l_s}^*)_i = 0, \forall i$. So, if $l_i < x_i^* < u_i$, then $(g_{l_s}^*)_i = 0$. Suppose $x_i^* = l_i$. Then we must have $(g_{l_s}^*)_i \ge 0$. Since otherwise, $(g_{l_s}^*)_i < 0$ and by (3.60), $x_i^{k+1} > x_i^k$ for all k sufficiently large. Then, $\lim x_i^k \ne l_i$, which is a contradiction to that $\lim x_i^k = x_i^* = l_i$. Similarly, we can show that if $x_i^* = u_i$, then $(g_{l_s}^*)_i \le 0$. ■

Condition (3.60) is crucial to the first-order convergence result; therefore, it is important to realize it can be satisfied with the scaled gradient direction, p_g^k. This we prove next.

LEMMA 3.8. *The vector p_g^k satisfies* **(CD2)**, **(CD3)** *and (3.60) $\forall k$ sufficiently large, where p_g^k is defined by (2.35).*

Proof. **(CD2)** and **(CD3)** are obviously satisfied. To prove (3.60), note that $(g_{l_s}^*)_i \ne 0, \forall i \in \mathcal{A}^*$. So when k is sufficiently large, $sign((g_{l_s}^k)_i) = sign((g_{l_s}^*)_i)$, and by (2.39) and (3.17), $\mu^k < 0$. Hence, by (2.35), we have

$$sign((p_g^k)_i) = sign(\mu^k \frac{\tilde{v}_i^k (g_{l_s}^k)_i}{\|D^k g_{l_s}^k\|}) = -sign((g_{l_s}^k)_i) = -sign((g_{l_s}^*)_i), \ \forall i \in \mathcal{A}^*. \tag{3.61}$$

[3] i.e., x^* satisfies condition (1.4).

∎

One implication of the previous two results is that algorithm **Interior-Newton$_1$** guarantees first-order convergence if we replace s_k with the scaled gradient direction whenever (3.60) is not satisfied.

Now, we turn to the second-order optimality condition. The following lemma reveals the relation between the definiteness of H in $\mathcal{N}$ and that of $\bar{Z}^T\bar{M}\bar{Z}$.

LEMMA 3.9. *Suppose x^* satisfies (3.21). Let*

$$\hat{M}^* \stackrel{\text{def}}{=} (\bar{Z}^*)^T\bar{M}^*\bar{Z}^* = (\bar{Z}^*)^T(D^*HD^* + |G^*_{ls}|)\bar{Z}^*.$$

Then

(i) $H \geq 0$ *in* $\mathcal{N}^*$ *if and only if* $\hat{M}^* \geq 0$.
(ii) $H > 0$ *in* $\mathcal{N}^*$ *if and only if* $\hat{M}^* > 0$.

Proof. We will prove (ii). (i) can be proved similarly. Assume first that $H > 0$ in $\mathcal{N}^*$. Let $s \in \mathcal{R}^{n-m}$ and $s \neq 0$. By **(AS2)**, AD^* has full row rank. So $\bar{Z}^* \in \mathcal{R}^{n\times(n-m)}$ and $\bar{Z}^*s \neq 0$. Let

$$\nu \stackrel{\text{def}}{=} s^T\hat{M}^*s = (D^*\bar{Z}^*s)^TH(D^*\bar{Z}^*s) + (\bar{Z}^*s)^T|G^*_{ls}|\bar{Z}^*s.$$

Note that $AD^*\bar{Z}^* = 0$ and $(D^*\bar{Z}^*s)_i = 0$, $\forall i \in \mathcal{A}^*$, we have $D^*\bar{Z}^*s \in \mathcal{N}^*$. So $(D^*\bar{Z}^*s)^TH(D^*\bar{Z}^*s) \geq 0$. Hence $\nu \geq 0$. If $D^*\bar{Z}^*s \neq 0$, then $\nu > 0$, we are all set. Otherwise, $(\bar{Z}^*s)_i = 0$, $\forall i \in \mathcal{A}^{*^c}$. Since $\bar{Z}^*s \neq 0$, $\exists i \in \mathcal{A}^*$ such that $(\bar{Z}^*s)_i \neq 0$. But by **(AS3)**, $(g^*_{ls})_i \neq 0$, $\forall i \in \mathcal{A}^*$. So $\nu > 0$, and therefore, $\hat{M}^* > 0$.

Now assume that $\hat{M}^* > 0$. Let $x \in \mathcal{N}^*$ and $x \neq 0$. Define $\tilde{x} \in \mathcal{R}^n$ as follows:

$$\tilde{x}_i = \begin{cases} \frac{x_i}{D^*_{ii}}, & \forall i \in \mathcal{A}^{*^c} \\ 0, & \text{otherwise.} \end{cases}$$

Then $x = D^*\tilde{x}$. Since $AD^*\tilde{x} = Ax = 0$, $\exists\hat{x}$ such that $\tilde{x} = \bar{Z}^*\hat{x}$. Note that $x \in \mathcal{N}^*$ and $x \neq 0$, we have $\tilde{x} \neq 0$ and $\hat{x} \neq 0$. Hence

$$0 < \hat{x}^T\hat{M}^*\hat{x} = \hat{x}^T(\bar{Z}^*)^TD^*HD^*\bar{Z}^*\hat{x} + \hat{x}^T(\bar{Z}^*)^T|G^*_{ls}|\bar{Z}^*\hat{x} = x^THx,$$

which implies that $H > 0$ in $\mathcal{N}^*$. ∎

Before proceeding to the second-order optimality conditions, we justify the fact that θ defined by (2.23) is a good measure of optimality.

LEMMA 3.10. *Let θ be defined by (2.23). Then $\theta = 0$ if and only if x satisfies the first-order and second-order optimality conditions.*

Proof. We first assume that $\theta = 0$. Then $v. * g = 0$ which implies that x satisfies the first-order condition. In addition, $\psi(p) = 0$. So by (2.14) and (2.15), we have

$$\frac{1}{2}\bar{p}^T\bar{Z}^T\bar{M}\bar{Z}\bar{p} + \lambda\|\bar{p}\|^2 = 0. \tag{3.62}$$

By (2.15) and (2.16), $\bar{p}^T\bar{Z}^T\bar{g} \leq 0$. So $\bar{\psi}(\bar{p}) = \psi(p) = 0$ yields

$$\frac{1}{2}\bar{p}^T\bar{Z}^T\bar{M}\bar{Z}\bar{p} \geq 0. \tag{3.63}$$

Hence,

$$\lambda \|\bar{p}\|^2 = 0. \tag{3.64}$$

Then, by (2.17) and $\lambda \geq 0$, $\lambda = 0$. Therefore, by (2.16),

$$\bar{Z}^T \bar{M} \bar{Z} \geq 0, \tag{3.65}$$

which implies that x satisfies the second-order condition by Lemma 3.9.

Now we assume that x satisfies the first-order and second-order conditions. Then first, $\|v. * g\| = 0$. Also, $\bar{g} = 0$. So (2.15) implies

$$\bar{p}^T \bar{Z}^T \bar{M} \bar{Z} \bar{p} + \lambda \|\bar{p}\|^2 = 0. \tag{3.66}$$

Then by Lemma 3.9 and the second-order conditions, we see that $\bar{p}^T \bar{Z}^T \bar{M} \bar{Z} \bar{p} \geq 0$ which by (3.66) implies that

$$\psi(p) = \bar{\psi}(\bar{p}) = \frac{1}{2} \bar{p}^T \bar{Z}^T \bar{M} \bar{Z} \bar{p} = 0. \tag{3.67}$$

■

The next result claims that if the point of convergence is a complementarity point, i.e., (3.21) is satisfied, then $\{\alpha\}$ defined by (2.27) is bounded away from zero.

LEMMA 3.11. *Suppose $\{w^k\}$ satisfies* **(CD1)**. *Suppose $\{x^k\}$ converges to a point x^* that satisfies (3.21). Then $\exists \alpha_0 > 0$ such that $\alpha^k \geq \alpha_0$, $\forall k$, where α^k is defined by (2.27).*

Proof. If the lemma is false, then $\exists$ a subsequence $\{k_j\}$ such that

$$\alpha^{k_j} \longrightarrow 0. \tag{3.68}$$

Therefore, by (2.27), $\beta^{k_j} \longrightarrow 0$. Since n is finite, from the definition of β, we may (without loss of generality) assume that $\beta^{k_j} = \frac{u_1 - x_1^{k_j}}{(p^{k_j})_1}$. Note that $\|p^k\| \leq \Delta_u \|D^k\|$ is bounded by (2.13) and the convergence of $\{x^k\}$, so $u_1 - x_1^{k_j} \longrightarrow 0$ and when k_j is sufficiently large, $\tilde{v}_1^{k_j} = u_1 - x_1^{k_j} \longrightarrow 0$. Hence, $\tilde{v}_1^* = 0$ and by **(AS3)**, $|(g_{ls}^*)_1| > 0$. But by (2.19),

$$D^2 H p + |G| p + \lambda p + D^2 g + D^2 A^T p_w = 0, \tag{3.69}$$

and so

$$-\frac{|g_1^{k_j}| + \lambda^{k_j}}{(H p^{k_j})_1 + g_1^{k_j} + (A^T p_w^{k_j})_1} = \frac{\tilde{v}_1^{k_j}}{(p^{k_j})_1} \longrightarrow 0. \tag{3.70}$$

Therefore, by the fact that $g_1^{k_j} \longrightarrow (g_{ls}^*)_1$, we must have

$$|(H p^{k_j})_1 + g_1^{k_j} + (A^T p_w^{k_j})_1| \longrightarrow \infty, \tag{3.71}$$

which is impossible since by **(AS2)**, **(CD1)**, (2.20) and the convergence of $\{x^k\}$, $\|H p^k + g^k + A^T p_w^k\|$ is bounded. ■

Guaranteeing a reduction in ψ^k which is a fraction of the reduction obtained by the truncated trust region step, $\alpha^k p^k$, where α^k and p^k are defined by (2.27) and (2.13), allows us to establish convergence to a second-order point. This is the reason for the following condition.

$$
\textbf{(CD4)} \quad \left\{ \begin{array}{c} \text{We say that a vector } \xi \text{ satisfies } \textbf{(CD4)} \text{ if} \\ \psi^k(\xi) \leq \gamma\, \psi^k(\alpha^k p^k), \\ \text{where } \alpha^k \text{ and } p^k \text{ are defined by (2.27) and (2.13).} \end{array} \right.
$$

THEOREM 3.12. *Suppose $\{w^k\}$ satisfies* **(CD1)** *and s^k satisfies* **(CD4)** *for all k sufficiently large. Suppose $\{x^k\}$ converges to a point x^* that satisfies (3.21). Then x^* satisfies (1.5).*

Proof. By (3.2), **(CD4)** and (3.3), for all k sufficiently large,

$$
q(x^k) - q(x^{k+1}) \geq -\gamma\, \psi^k(\alpha^k p^k) \geq -\gamma \frac{\alpha^k}{2} \lambda^k \|\bar{p}^k\| \longrightarrow 0. \tag{3.72}
$$

By Lemma 3.11, we see that $\lambda^k \|\bar{p}^k\| \longrightarrow 0$, and by (2.17),

$$
\lambda^k \longrightarrow 0. \tag{3.73}
$$

Hence, by (2.16),

$$
(\bar{Z}^*)^T \bar{M}^* \bar{Z}^* \geq 0. \tag{3.74}
$$

Therefore, (1.5) holds at x^* by Lemma 3.9. ∎

To sum up, we have the following theorem:

THEOREM 3.13. *Suppose $\{w^k\}$ satisfies* **(CD1)** *and s^k satisfies* **(CD2)**, **(CD3)**, **(CD4)**, *and (3.60) for all k sufficiently large. Then the sequence $\{x^k\}$ generated by Algorithm* **Interior-Newton$_1$** *converges to a point x^* which satisfies the first-order and second-order necessary optimality conditions. Further more, if $H > 0$ in $\mathcal{N}^*$, then x^* is a solution to (1.1).*

As pointed out before, **(CD1)** can be satisfied by letting $w^k = w^k_{ls}$. We now show how **(CD2)**, **(CD3)**, **(CD4)**, and (3.60) can be satisfied simultaneously. First, we define a parameter as follows: let $0 < \tau < 1$ be given, let $\xi, \eta \in \mathcal{R}^n$,

$$
\sigma = \sigma^k(\xi, \eta) \stackrel{\text{def}}{=} \left\{ \begin{array}{ll} 1, & \text{if } \psi^k(\xi) \leq \tau\, \psi^k(\eta), \\ 0, & \text{otherwise.} \end{array} \right. \tag{3.75}
$$

Clearly, if α^k, p^k, α^k_g and p^k_g are defined by (2.27), (2.13), (2.41) and (2.35), and $s^k = \sigma^k \alpha^k p^k + (1-\sigma^k)\alpha^k_g p^k_g$, **(CD2)**, **(CD3)** and **(CD4)** are satisfied. Next, we show that (3.60) is also satisfied.

LEMMA 3.14. *Suppose $\{w^k\}$ satisfies* **(CD1)** *and s^k satisfies* **(CD2)**, **(CD3)**, *and* **(CD4)** *for all k sufficiently large. Then for all subsequences $\{p^{k_j}\}$ such that $p^{k_j} \longrightarrow 0$, we have*

$$
sign(p^{k_j}_i) = -sign((g^*_{ls})_i), \quad \forall i \in \mathcal{A}^*, \quad \forall k_j \text{ sufficiently large,} \tag{3.76}
$$

where p^{k_j} is defined by (2.13).

Proof. By Theorem 3.6 and Corollary 3.3, $\{x^k\}$ converges to x^* where x^* satisfies (3.21). Also, by **(AS2)**, **(CD1)**, (2.20) and the convergence of $\{x^k\}$, $\|Hp^k + g^k + A^T p_w^k\|$ is bounded, and by **(AS3)**, $|(g_{ls}^*)_i| > 0$, $\forall i \in \mathcal{A}^*$. So, by (3.69), $\exists C_2 > 0$ such that for all k sufficiently large,

$$(3.77) \qquad |p_i^k| \le \tilde{v}_i^k \frac{|(Hp^k)_i + g_i^k + (A^T p_w^k)_i|}{|g_i^k| + \lambda^k} \le C_2 \tilde{v}_i^k, \ \forall i \in \mathcal{A}^*,$$

and consequently, by (2.13),

$$|(\bar{Z}^k \bar{p}^k)_i| = \frac{|p_i^k|}{(\tilde{v}_i^k)^{\frac{1}{2}}} \le C_2 (\tilde{v}_i^k)^{\frac{1}{2}} \longrightarrow 0, \ \forall i \in \mathcal{A}^*.$$

So,

$$(3.78) \qquad |G^k| \bar{Z}^k \bar{p}^k \longrightarrow 0, \ \text{since } g_i^k \longrightarrow 0, \ \forall i \in \mathcal{A}^{*^c}.$$

Hence by the assumption that $D^{k_j} \bar{Z}^{k_j} \bar{p}^{k_j} = p^{k_j} \longrightarrow 0$, we have

$$(3.79) \qquad \bar{M}^{k_j} \bar{Z}^{k_j} \bar{p}^{k_j} = D^{k_j} H D^{k_j} \bar{Z}^{k_j} \bar{p}^{k_j} + |G^{k_j}| \bar{Z}^{k_j} \bar{p}^{k_j} \longrightarrow 0.$$

Therefore, by (2.20), **(AS2)**, (3.73), and the fact that $\bar{g}^k \longrightarrow 0$, we have

$$(3.80) \qquad p_w^{k_j} \longrightarrow 0.$$

Finally, using (3.69) again, for all k_j sufficiently large and $i \in \mathcal{A}^*$, we have

$$(3.81) \qquad \begin{aligned} sign(p_i^{k_j}) &= sign\left(-\tilde{v}_i^{k_j} \frac{(Hp^{k_j})_i + g_i^{k_j} + (A^T p_w^{k_j})_i}{|g_i^{k_j}| + \lambda^{k_j}}\right) \\ &= sign(-g_i^{k_j}) = -sign((g_{ls}^*)_i). \end{aligned}$$

■

THEOREM 3.15. *Suppose for all k sufficiently large, $w^k = w_{ls}^k$ and $s^k = \sigma^k \alpha^k p^k + (1 - \sigma^k) \alpha_g^k p_g^k$, where α^k, p^k, α_g^k and p_g^k are defined by (2.27), (2.13), (2.41) and (2.35). Then the sequence $\{x^k\}$ generated by Algorithm* **Interior-Newton$_1$** *converges to a point x^* which satisfies the first-order and second-order necessary optimality conditions. Further more, if $H > 0$ in $\mathcal{N}^*$, then x^* is a solution to (1.1).*

Proof. By Theorem 3.13, we only need to show that s^k satisfies (3.60) for all k sufficiently large. By (3.61), it suffices to show that

$$(3.82) \quad sign(p_i^k) = -sign((g_{ls}^*)_i), \ \forall i \in \mathcal{A}^*, \ \forall k \text{ sufficiently large, and } \sigma^k = 1.$$

In fact, for all k such that $\sigma^k = 1$, we have $s^k = \alpha^k p^k$. So by Lemma 3.11 and the convergence of $\{x^k\}$,

$$(3.83) \qquad \{p^k : \sigma^k = 1\} \longrightarrow 0.$$

Hence, by Lemma 3.14, we see that (3.82) holds. ■

We complete this section with two observations.

First, we remark on the computational cost of each iteration. If Algorithm **Interior-Newton$_1$** is implemented with $w^k = w_{ls}^k$ and $s^k = \sigma^k \alpha^k p^k + (1-\sigma^k)\alpha_g^k p_g^k$, then the main cost in each iteration is the QR-factorization of AD^k (to get $\bar{Z}^k$ and w^k) and a trust region subproblem solution. This is reasonably practical for small and medium-sized problems.

Second, we note that our choice of σ in (3.75) is just one of numerous possible ways to allow for an effective combination of the scaled gradient and trust region directions. Another possibility, for example, is to follow the "dogleg" idea, (e.g., [18]), a popular idea used in unconstrained minimization settings. Specifically, let

$$\sigma = \sigma^k(\xi, \eta) \stackrel{\text{def}}{=} \{\, 0 \le \sigma \le 1 \,:\, \phi^k(\sigma) \stackrel{\text{def}}{=} \psi^k(\sigma\xi + (1-\sigma)\eta) = \min \,\}. \tag{3.84}$$

4. Quadratic Convergence of Algorithm Interior-Newton$_1$.

In this section we consider the local convergence rate properties of algorithm **Interior-Newton$_1$**: we establish superlinear and quadratic convergence results.

Recap: For ease of reference we summarize here the global assumptions used throughout this section and the next. Recall that $\tilde{V}$ is a diagonal matrix, defined by (1.12), and $\mathcal{C}$ is the set defined by (3.13), i.e.,

$$\mathcal{C} \stackrel{\text{def}}{=} \{x \in \mathcal{F} \,:\, \tilde{v}.*(Hx + c + A^T w) = 0 \ \text{ for some } \ w \in \mathcal{R}^m\}. \tag{4.1}$$

The least-squares multipliers, w_{ls}, are defined in (3.9) and (3.10).

(AS1) The level set $\mathcal{L} = \{x \in \mathcal{F} \,:\, q(x) \le q(x^0)\}$ is compact.

(AS2) $A\tilde{V}A^T$ is nonsingular $\forall x \in \mathcal{L}$.

(AS3) $\forall x \in \mathcal{C}$, if $(Hx + c + A^T w_{ls})_i = 0$, then $\tilde{v}_i \ne 0$.

We also summarize the various conditions that will be referred to in different results to follow. Bear in mind that ultimately the conditions will be satisfied by appropriate and computable quantities.

(CD1) We say that $\{w^k\}$ satisfies **(CD1)** if $w^{k_j} \longrightarrow w_{ls}^*$ whenever $x^{k_j} \longrightarrow x^* \in \mathcal{C}$.

(CD2) $\left\{ \begin{array}{l} \text{A vector } \xi \text{ satisfies } \textbf{(CD2)} \text{ if} \\ \quad \psi^k(\xi) \le \gamma\, \psi^k(\alpha_g^k p_g^k), \\ \text{where } \alpha_g^k \text{ and } p_g^k \text{ are defined by (2.41) and (2.35).} \end{array} \right.$

(CD3) $\left\{ \begin{array}{l} \text{A vector } \xi \text{ satisfies } \textbf{(CD3)} \text{ if} \\ \quad \|(D^k)^{-1}\xi\| \le \gamma_0\, \Delta^k. \end{array} \right.$

(CD4) $\left\{ \begin{array}{l} \text{A vector } \xi \text{ satisfies } \textbf{(CD4)} \text{ if} \\ \quad \psi^k(\xi) \le \gamma\, \psi^k(\alpha^k p^k), \\ \text{where } \alpha^k \text{ and } p^k \text{ are defined by (2.27) and (2.13).} \end{array} \right.$

We assume that for all k sufficiently large, w^k satisfies **(CD1)**, s^k satisfies **(CD2)**,

(CD3), **(CD4)**, and (3.60). Therefore, by Theorem 3.13 and Lemma 3.10,

$$w^k \longrightarrow w^*_{ls}, \tag{4.2}$$

(4.3) $x^k \longrightarrow x^*$ satisfying the first-order and second-order optimality conditions,

and

$$\theta^k \longrightarrow 0, \tag{4.4}$$

where θ^k is defined by (2.23).

LEMMA 4.1. *Assume $H > 0$ in $\mathcal{N}^*$. Then $p^k \longrightarrow 0$, $p^k_w \longrightarrow 0$, $\lambda^k = 0$, and $(\bar{Z}^k)^T \bar{M}^k \bar{Z}^k > 0$ for all k sufficiently large, where p^k and p^k_w are defined by (2.13) and (2.20), λ^k is defined by (2.15).*

Proof. Since $H > 0$ in $\mathcal{N}^*$, by Lemma 3.9, $\hat{M}^* > 0$. So by (3.73), $\exists C_3 > 0$, such that for all k sufficiently large, $(\bar{Z}^k)^T \bar{M}^k \bar{Z}^k + \lambda^k I > 0$ and

$$\|((\bar{Z}^k)^T \bar{M}^k \bar{Z}^k + \lambda^k I)^{-1}\| \leq C_3. \tag{4.5}$$

Hence, from (2.15) and the fact that $\bar{g}^k \longrightarrow 0$, we have

$$\bar{p}^k \longrightarrow 0. \tag{4.6}$$

Therefore,

$$p^k = D^k \bar{Z}^k \bar{p}^k \longrightarrow 0, \tag{4.7}$$

and by (2.17),

$$\lambda^k = 0, \ \forall k \text{ sufficiently large}. \tag{4.8}$$

So,

$$(\bar{Z}^k)^T \bar{M}^k \bar{Z}^k > 0, \ \forall k \text{ sufficiently large}. \tag{4.9}$$

Finally, similar to (3.80), we have

$$p^k_w \longrightarrow 0. \tag{4.10}$$

∎

LEMMA 4.2. *Suppose $H > 0$ in $\mathcal{N}^*$. Then $\alpha^k \longrightarrow 1$, where α^k is defined by (2.27).*

Proof. $\forall i \in \mathcal{A}^{*^c}$, $\tilde{v}^*_i > 0$. So by (4.7),

$$BR^k_i = \max\{\frac{l_i - x^k_i}{p^k_i}, \frac{u_i - x^k_i}{p^k_i}\} \longrightarrow \infty, \ \forall i \in \mathcal{A}^{*^c}. \tag{4.11}$$

Recall that $(g^*_{ls})_i \neq 0$, $\forall i \in \mathcal{A}^*$. Without loss of generality, assume that $(g^*_{ls})_i > 0$. Then $x^*_i = l_i$ and similar to (3.81), when k is sufficiently large, $p^k_i < 0$. By considering (3.69), (4.7), (4.8), and (4.10), we see that

$$BR^k_i = \frac{l_i - x^k_i}{p^k_i} = -\frac{\tilde{v}^k_i}{p^k_i} = \frac{|g^k_i| + \lambda^k}{(Hp^k)_i + g^k_i + (A^T p^k_w)_i} \longrightarrow 1, \ \forall i \in \mathcal{A}^*. \tag{4.12}$$

Therefore, by (2.25), (2.26), (2.27), and (4.4), we have $\beta^k \longrightarrow 1$, $\rho^k \longrightarrow 1$, and $\alpha^k \longrightarrow 1$. ∎

Next we cite a standard result (see, e.g., [7]), used subsequently.

THEOREM 4.3. *Let $\mathcal{D} \subseteq \Re^n$ be an open convex set, $F : \Re^n \longrightarrow \Re^n$. $y^* \in \mathcal{D}$, $F(y^*) = 0$, $\nabla F(y^*)$ nonsingular, and ∇F is Lipschitz continuous at y^* in $\mathcal{D}$. Let $\{T_k\}$ be a sequence of nonsingular matrices in $\Re^{n\times n}$, and suppose for some $y^0 \in \mathcal{D}$ that the sequence of points generated by $y^{k+1} = y^k - T_k^{-1}F(y^k)$ remains in $\mathcal{D}$, and satisfies $y^k \neq y^*$ for all k, and $y^k \longrightarrow y^*$.*
If $\|T_k - \nabla F(y^*)\| \longrightarrow 0$, then $\{y^k\}$ converges superlinearly to y^*.
If $\|T_k - \nabla F(y^*)\| = O(\|y^k - y^*\|)$, then $\{y^k\}$ converges quadratically to y^*.

Theorem 4.3 can not be applied directly to $\tilde{F}$ or F since these systems are not differentiable at some points. Instead, we apply Theorem 4.3 to the following function:

$$\hat{F}(x,w) \stackrel{\text{def}}{=} \begin{bmatrix} \hat{v}.*g \\ Ax-b \end{bmatrix} : \mathcal{R}^{m+n} \longrightarrow \mathcal{R}^{m+n}, \tag{4.13}$$

where

$$\hat{v}(x)_i \stackrel{\text{def}}{=} \begin{cases} \frac{l_i+u_i}{2}, & \text{if } x_i^* = \frac{l_i+u_i}{2} \\ \tilde{v}_i, & \text{otherwise.} \end{cases} \tag{4.14}$$

It is clear that $\hat{v}(x^*) = \tilde{v}(x^*)$. So

$$\hat{F}(x^*, w_{ls}^*) = 0. \tag{4.15}$$

Also, $\nabla\hat{F}$ exists and continuous everywhere in $\mathcal{R}^{m+n}$. Since

$$(g_{ls})_i^* = 0, \;\; \forall i \in \mathcal{A}^{*^c},$$
$$(\nabla\hat{v}^*)_{ii} = sign((g_{ls})_i^*), \;\forall i \in \mathcal{A}^*,$$

we have, in a neighborhood of (x^*, w_{ls}^*),

$$\nabla\hat{F}(x,w) = \begin{bmatrix} \nabla\hat{v}G + \hat{V}H & \hat{V}A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} |G| + \hat{V}H & \hat{V}A^T \\ A & 0 \end{bmatrix}. \tag{4.16}$$

LEMMA 4.4. *$\nabla\hat{F}(x^*, w_{ls}^*)$ is nonsingular.*
Proof. Let $\nabla\hat{F}^*s = 0$, where $s = (s_x, s_w)$ has appropriate size. Then by (4.16),

$$As_x = 0 \tag{4.17}$$
$$(|G_{ls}^*| + \hat{V}^*H)s_x + \hat{V}^*A^Ts_w = 0. \tag{4.18}$$

It follows from (4.18) that $\hat{v}_{\mathcal{A}^*}^* = 0$, we have

$$|(g_{ls}^*)_{\mathcal{A}^*}|.*(s_x)_{\mathcal{A}^*} = \hat{v}_{\mathcal{A}^*}^*.*(Hs_x)_{\mathcal{A}^*} + \hat{v}_{\mathcal{A}^*}^*.*(A^Ts_w)_{\mathcal{A}^*} = 0. \tag{4.19}$$

But $(g_{ls}^*)_i \neq 0$, $\forall i \in \mathcal{A}^*$, so $(s_x)_{\mathcal{A}^*} = 0$ and by (4.17), we see that

$$s_x \in \mathcal{N}^*. \tag{4.20}$$

Also from (4.18), and that $(g_{ls}^*)_{\mathcal{A}^{*c}} = 0$, $\hat{v}_i^* \neq 0$, $\forall i \in \mathcal{A}^{*^c}$, we have $(Hs_x)_{\mathcal{A}^{*c}} + (A^T s_w)_{\mathcal{A}^{*c}} = 0$. Thus

$$(4.21)\quad s_x^T H s_x = s_x^T H s_x + s_x^T(A^T s_w) = (s_x)_{\mathcal{A}^{*c}}^T (H s_x)_{\mathcal{A}^{*c}} + (s_x)_{\mathcal{A}^{*c}}^T (A^T s_w)_{\mathcal{A}^{*c}} = 0.$$

By the assumption that $H > 0$ in $\mathcal{N}^*$, we have $s_x = 0$. Then from (4.18) and (A.3), $s_w = 0$. So $s = 0$, which implies that $\nabla \hat{F}^*$ is nonsingular. ∎

LEMMA 4.5. *$\exists \delta > 0$ and a neighborhood $B_\delta(x^*, w_{ls}^*)$ of (x^*, w_{ls}^*), such that $\nabla \hat{F}$ is Lipschitz continuous at (x^*, w_{ls}^*) in $B_\delta(x^*, w_{ls}^*)$, and $\forall (x, w) \in B_\delta(x^*, w_{ls}^*)$, $\nabla \hat{F}$ is nonsingular.*

Proof. The existence of $(\nabla \hat{F}(x, w))^{-1}$ in $B_\delta(x^*, w_{ls}^*)$ is obvious by Lemma 4.4. For the Lipschitz continuity, $\forall (x, w) \in B_\delta(x^*, w_{ls}^*)$, $\exists C_6 > 0$, such that

$$(4.22)\quad \begin{aligned} & \|\nabla \hat{F}(x, w) - \nabla \hat{F}(x^*, w_{ls}^*)\| \\ = & \left\| \begin{bmatrix} \nabla \hat{v} G + \hat{V} H - \nabla \hat{v}^* G_{ls}^* - \hat{V}^* H & (\hat{V} - \hat{V}^*) A^T \\ 0 & 0 \end{bmatrix} \right\| \\ \leq & \; C_6 \left\| \begin{bmatrix} x - x^* \\ w - w_{ls}^* \end{bmatrix} \right\|. \end{aligned}$$

∎

It is not clear how to directly apply Theorem 4.3 to the sequence $\{(x^k, w^k)\}$ since $\{w^k\}$ is not updated in the form of $w^{k+1} = w^k + p_w^k$. Therefore, to establish the convergence rates of $\{(x^k, w^k)\}$ and $\{x^k\}$, intermediate steps need be inspected. Specifically, we consider the sequence $\{(x^k, \hat{w}^k)\}$ where $\{\hat{w}^k\}$ is defined by

$$(4.23)\quad \hat{w}^k = w^{k-1} + p_w^{k-1}, \; \forall k = 1, 2, \ldots,$$

and $\{p_w^k\}$ is defined by (2.20). By considering (4.10) and **(CD1)**, we see that

$$(4.24)\quad \hat{w}^k \longrightarrow w_{ls}^*.$$

By (2.19), note that $\lambda^k = 0$ for all k sufficiently large, we have

$$(4.25)\quad (H + D^{-2}|G|)p + A^T p_w = -g.$$

So, if for $k = 1, 2, \ldots$, we let

$$\begin{aligned} (4.26)\quad \hat{p}_w^k &= \hat{w}^{k+1} - \hat{w}^k, \\ (4.27)\quad \hat{g}^k &= Hx^k + c + A^T \hat{w}^k, \end{aligned}$$

then

$$(4.28)\quad (H + D^{-2}|G|)p + A^T \hat{p}_w = -\hat{g}.$$

Define

$$(4.29)\quad T_\alpha \stackrel{\text{def}}{=} \begin{bmatrix} \frac{1}{\alpha}(\hat{V} D^{-2}|G| + \hat{V} H) & \hat{V} A^T \\ \frac{1}{\alpha} A & 0 \end{bmatrix}.$$

Since $\hat{V}^k(D^k)^{-2}|G^k| \longrightarrow |G^*_{ls}|$, by Lemma 4.2, we have

$$T^k_\alpha \longrightarrow \nabla \hat{F}^*. \tag{4.30}$$

So $\forall k$ sufficiently large, T^k_α is nonsingular and

$$(\alpha^k p^k, \hat{p}^k_w) = -(T^k_\alpha)^{-1}\hat{F}(x^k, \hat{w}^k). \tag{4.31}$$

Hence,

$$(x^{k+1}, \hat{w}^{k+1}) = (x^k, \hat{w}^k) - (T^k_\alpha)^{-1}\hat{F}(x^k, \hat{w}^k). \tag{4.32}$$

Using Theorem 4.3 and (4.30), we have the following theorem:

THEOREM 4.6. *Suppose $H > 0$ in $\mathcal{N}^*$. Suppose for all k sufficiently large, $s^k = \alpha^k p^k$, where α^k and p^k are defined by (2.27) and (2.13). Then $(x^k, \hat{w}^k)$ converges to (x^*, w^*_{ls}) superlinearly.*

Next, we show that $\{(x^k, \hat{w}^k)\}$ converges quadratically to (x^*, w^*_{ls}).

LEMMA 4.7. *Suppose $H > 0$ in $\mathcal{N}^*$. Suppose for all k sufficiently large, $w^k = w^k_{ls}$ and $s^k = \alpha^k p^k$, where α^k and p^k are defined by (2.27) and (2.13). Then $|1 - \alpha^k| = O(\|x^k - x^*\|)$.*

Proof. First, for all k sufficiently large,

$$\begin{aligned} w^*_{ls} - w^k &= w^*_{ls} - w^k_{ls} \\ &= (A\tilde{V}^*A^T)^{-1}A\tilde{V}^*(Hx^* + c) - (A\tilde{V}^kA^T)^{-1}A\tilde{V}^k(Hx^k + c), \end{aligned} \tag{4.33}$$

By (3.5), it is easy to verify that

$$\|w^k - w^*_{ls}\| = O(\|x^k - x^*\|). \tag{4.34}$$

Consequently,

$$\|\tilde{v}^k. * g^k\| = O(\|x^k - x^*\|). \tag{4.35}$$

By (2.13) and (2.19), we have

$$\begin{bmatrix} |G^k| + \tilde{V}^k H & \tilde{V}^k A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} p^k \\ p^k_w \end{bmatrix} = \begin{bmatrix} -\tilde{v}^k. * g^k \\ 0 \end{bmatrix}. \tag{4.36}$$

It is easy to see that the limit of the coefficient matrix is $\nabla \hat{F}^*$. So, by (4.35),

$$\|p^k\| + \|p^k_w\| = O(\|x^k - x^*\|), \tag{4.37}$$

$$\psi^k(p^k) = O(\|x^k - x^*\|), \tag{4.38}$$

$$\theta^k = O(\|x^k - x^*\|), \text{ and } 1 - \rho^k = O(\|x^k - x^*\|). \tag{4.39}$$

Meanwhile, similar to (4.12), $\forall i \in \mathcal{A}^*$ and all k sufficiently large,

$$|1 - BR^k_i| = |\frac{(Hp^k)_i + (A^Tp^k_w)_i}{(Hp^k)_i + g^k_i + (A^Tp^k_w)_i}| = O(\|x^k - x^*\|). \tag{4.40}$$

So,

$$|1-\beta^k| = O(\|x^k - x^*\|), \tag{4.41}$$

and therefore, by (4.39),

$$|1-\alpha^k| = O(\|x^k - x^*\|). \tag{4.42}$$

■

THEOREM 4.8. *The sequence $\{(x^k, \hat{w}^k)\}$ converges to (x^*, w^*_{ls}) quadratically under the assumptions of Lemma 4.7.*

Proof. We show that

$$\|T^k_\alpha - \nabla \hat{F}^*\| = O(\|x^k - x^*\|). \tag{4.43}$$

Then by Theorem 4.3, the result follows. In fact, let $\|\cdot\|_F$ denote the Frobenius norm, we have

$$\begin{aligned} \|T^k_\alpha - \nabla \hat{F}^*\|_F &= \|\frac{1}{\alpha^k}\hat{V}^k(D^k)^{-2}|G^k| - |G^*_{ls}|\,\|_F + \|(\frac{1}{\alpha^k}\hat{V}^k - \hat{V}^*)\,H\|_F \\ &\quad + \|(\hat{V}^k - \hat{V}^*)\,A^T\|_F + \|(\frac{1}{\alpha^k} - 1)\,A\|_F. \end{aligned} \tag{4.44}$$

It is not hard to show that

$$\|\hat{V}^k(D^k)^{-2}|G^k| - |G^*_{ls}|\,\|_F + \|\hat{V}^k - \hat{V}^*\|_F = O(\|x^k - x^*\|). \tag{4.45}$$

So by (4.42), and the equivalence of the norms, (4.43) holds. ■

Next we establish the convergence rate of the sequences $\{x^k\}$ and $\{(x^k, w^k)\}$. To do this, we need the following lemma which can be regarded as a complement to Theorem 4.3. Its proof is similar to that of Theorem 8.2.4 in [7].

LEMMA 4.9. *Suppose in Theorem 4.3, there is a partition $y = (y_1, y_2)$ and $\|T_k - \nabla F(y^*)\| = O(\|y^k_1 - y^*_1\|)$. Then $\exists C > 0$ such that for all k sufficiently large,*

$$\|y^{k+1}_1 - y^*_1\| \le C\,\|y^{k+1} - y^k\|\,\|y^k_1 - y^*_1\|. \tag{4.46}$$

THEOREM 4.10. *Suppose $H > 0$ in $\mathcal{N}^*$. Suppose for all k sufficiently large, $w^k = w^k_{ls}$ and $s^k = \alpha^k p^k$, where α^k and p^k are defined by (2.27) and (2.13). Then $\exists C > 0$ such that for all k sufficiently large,*

$$\|x^{k+1} - x^*\| \le C\,\|x^{k-1} - x^*\|\,\|x^k - x^*\|, \tag{4.47}$$

$$\begin{aligned} &\|(x^{k+1}, w^{k+1}) - (x^*, w^*_{ls})\| \\ &\le C\,\|(x^{k-1}, w^{k-1}) - (x^*, w^*_{ls})\|\,\|(x^k, w^k) - (x^*, w^*_{ls})\|. \end{aligned} \tag{4.48}$$

Proof. By (4.43) and Lemma 4.9, $\exists C > 0$ such that for all k sufficiently large,

$$\|x^{k+1} - x^*\| \le C\,\|(x^{k+1}, \hat{w}^{k+1}) - (x^k, \hat{w}^k)\|\,\|x^k - x^*\|. \tag{4.49}$$

Therefore, (4.47) and (4.48) follow by (4.23), (4.34), and (4.37). ∎

Theorem 4.10 shows that the convergence rates of $\{x^k\}$ to x^* and $\{(x^k, w^k)\}$ to (x^*, w_{ls}^*) are almost quadratic, stronger than both superlinear and 2-step quadratic.

The last result of this section shows that $s^k = \alpha^k p^k$, where p_k is given by (2.13) and α_k is given by (2.27), will ultimately hold for all k. Therefore, the convergence rate results mentioned above will apply to algorithm **Interior-Newton$_1$**.

THEOREM 4.11. *Suppose $H > 0$ in $\mathcal{N}^*$. Suppose for all k sufficiently large, $s^k = \sigma^k \alpha^k p^k + (1 - \sigma^k)\alpha_g^k p_g^k$, where σ^k is defined by (3.75) and α^k, p^k, α_g^k and p_g^k are defined by (2.27), (2.13), (2.41) and (2.35). Then $s^k = \alpha^k p^k$ for all k sufficiently large.*

Proof. By (3.75), We need to show that

$$\psi^k(\alpha^k p^k) \leq \tau\, \psi^k(\alpha_g^k p_g^k), \quad \forall k \text{ sufficiently large.} \tag{4.50}$$

In fact, by (2.15) and (4.8), we have

$$\psi^k(\alpha^k p^k) = \alpha^k(2 - \alpha^k)\bar{\psi}^k(\bar{p}^k). \tag{4.51}$$

On the other hand, $\|(D^k)^{-1}\alpha_g^k p_g^k\| \leq \Delta^k$, so

$$\psi^k(\alpha_g^k p_g^k) \geq \psi^k(p^k) = \bar{\psi}^k(\bar{p}^k). \tag{4.52}$$

Therefore, by Lemma 4.2,

$$\frac{\psi^k(\alpha^k p^k)}{\psi^k(\alpha_g^k p_g^k)} = \frac{\alpha^k(2 - \alpha^k)\bar{\psi}^k(\bar{p}^k)}{\psi^k(\alpha_g^k p_g^k)} \geq \alpha^k(2 - \alpha^k) \longrightarrow 1. \tag{4.53}$$

Since $\tau < 1$, we see that (4.50) is true. ∎

5. An Accelerated Version of Algorithm Interior-Newton and Numerical Experiments. We have shown above that Algorithm **Interior-Newton$_1$** has strong convergence properties. However, this version of **Interior-Newton** uses $D^k = \tilde{V}^{\frac{1}{2}}$ which is based on the nonlinear system (1.17). System (1.17) is weaker than (1.16). In particular, (1.16) includes the "sign condition" for optimality, whereas (1.17) is derived solely from feasibility and complementary slackness conditions. Therefore, an algorithm based on (1.16) would be expected to outperform a similar algorithm based on the weaker system (1.17): non-optimal points, satisfying complementary slackness and feasibility, are attractors for (1.17) but not for (1.16).

This logic leads, apparently, to the use of $D^k = |V^k|^{\frac{1}{2}}$ in algorithm **Interior-Newton**. Unfortunately, the unbridled use of $D^k = |V^k|^{\frac{1}{2}}$ does not appear to yield a convergent process. Nevertheless, computational performance using $D^k = |V^k|^{\frac{1}{2}}$ often yields a significant improvement over the use of $D^k = \tilde{V}^{\frac{1}{2}}$. Consequently, we have developed an algorithm that mixes the use of these two different scaling matrices. We use $D^k = |V^k|^{\frac{1}{2}}$ when progress is good – in most cases this scaling accelerates progress toward a neighborhood of the solution. If progress is weak, $D^k = \tilde{V}^{\frac{1}{2}}$ is used. In this fashion impressive computational performance is achieved whilst maintaining the strong global convergence properties of algorithm **Interior-Newton$_1$**. The second-order convergence rate is maintained. We denote this algorithm, **Interior-Newton$_2$**. Specifically, algorithm **Interior-Newton$_2$** is defined to be **Interior-Newton** with Step 1 and Step 2 being described below.

Let $0 < \tau_1, \tau_2, \tau_3, \tau_4, \tau_5 < 1$, and $\gamma_1, \gamma_2, \Delta_l, \Delta_u > 0$ be given. For each k, let

$$\Delta^k \in [\Delta_l, \Delta_u]. \tag{5.1}$$

Step 1. First, select a trial D^k:

$D^0 = (\tilde{V}^0)^{\frac{1}{2}}$, and for $k \geq 1$,

If $\frac{\psi^{k-1}(\alpha^{k-1}p^{k-1})}{\psi^{k-1}(p^{k-1})} \leq \min(\tau_3, \tau_4\theta^{k-1})$ or $\frac{\psi^{k-1}(\alpha^{k-1}p^{k-1})}{\psi^{k-1}(\alpha_g^{k-1}p_g^{k-1})} \leq \tau_5$

$$D^k = (\tilde{V}^k)^{\frac{1}{2}}; \tag{5.2}$$

else

$$\check{g}^k = Hx^k + c + A^T w^{k-1}; \tag{5.3}$$

$$\check{v}^k = \mathcal{M}_{x^k}(\check{g}^k); \tag{5.4}$$

$$D^k = |\check{V}^k|^{\frac{1}{2}}; \tag{5.5}$$

end.

Then, perform an acceptance test to determine D^k:

If $D^k \neq (\tilde{V}^k)^{\frac{1}{2}}$

$$\check{t}^k = \|\check{v}^k .* \check{g}^k\| \,/\, (1 + \|\check{v}^k .* \check{g}^k\|); \tag{5.6}$$

compute w^k, $\bar{Z}^k$; e.g., by (2.33);
and compute p_g^k; e.g., by (2.35);
if $\psi^k(\alpha_g^k p_g^k) \,/\, \psi^k(p_g^k) \leq \min(\tau_3, \tau_4 \check{t}^k)$
$D^k = (\tilde{V}^k)^{\frac{1}{2}}$;
end
end

Finally:

If $D^k = (\tilde{V}^k)^{\frac{1}{2}}$
compute w^k, $\bar{Z}^k$; e.g., by (2.33);
and compute p_g^k; e.g., by (2.35);
end

Step 2 First, compute p^k; e.g., by (2.13); then compute s^k such that

$$\psi^k(s^k) \leq \min(\gamma_1\psi^k(\alpha^k p^k), \gamma_2\psi^k(\alpha_g^k p_g^k)). \tag{5.7}$$

In the Appendix, we verify that algorithm **Interior-Newton$_2$** has the same strong convergence properties as algorithm **Interior-Newton$_1$**. Here we make a few remarks before discussing our computational experiments.

1. As we show in the Appendix, due to the acceptance test in Step 1 for safeguarding, the convergence of $\{x^k\}$ is independent of the choice between $D^k = |\check{V}^k|^{\frac{1}{2}}$ and $D^k = (\tilde{V}^k)^{\frac{1}{2}}$. However, a proper choice is important for improved practical performance: the rule we specify has done well in our numerical experiments.
2. Note that the computation of $\check{g}^k$ is based on x^k and w^{k-1} since w^k is not available yet at this stage. (We could compute $\check{g}^k$ based on x^k and w_{ls}^k but an additional QR-factorization would be required.)
3. We show in the Appendix that, ultimately, only one computation is needed for w^k, $\bar{Z}^k$ and p_g^k in each iteration; therefore, ultimately there is no extra cost for the acceptance test.
4. To satisfy (5.7), we compute $s^k = \sigma^k \alpha^k p^k + (1-\sigma^k)\alpha_g^k p_g^k$, where σ^k is defined by (3.75).

We have implemented our algorithms in Matlab 4.0 and conducted some preliminary testing to investigate the practical viability of our approach. The experiments were performed on a Sun (Sparc) workstation. In the remainder of this section we present and discuss our preliminary numerical results.

Problem Generating: The Moré/Toraldo [15] QP-generator was adapted to generate the test problems. All the problems reported here are dense, and of moderate size. For positive definite problems the single global solution, with prescribed characteristics, was generated. For indefinite problems, the local minimizer determined by the algorithm(s) may have little relationship with point generated (with prescribed characteristics) by the test problem generator: indefinite problems have many local minimizers.

Starting and Stopping: For each test problem, a feasible starting point was found using a feasible point determination algorithm [13]. This procedure does not involve the matrix H nor the vector c; therefore, the feasible starting point can be considered somewhat arbitrary with regard to problem (1.1).

As usual, choosing a robust stopping criteria is not easy. We have used stopping criteria based on three computations: the relative difference in the objective function values of two successive iterations, the size of α, and the size of θ which is defined by (2.23). We terminate the iteration if:

$$\left\{ \begin{array}{c} q(x^k) - q(x^{k+1}) \leq tol * (1 + |q(x^k)|) \\ \text{and} \\ \alpha^k \geq 0.1, \end{array} \right. \tag{5.8}$$

or

$$\theta^k \leq tol, \tag{5.9}$$

or

$$k = 100. \tag{5.10}$$

Criterion (5.9) is reasonable since we have proven that $\theta = 0$ if and only if the first-order and second-order necessary optimality conditions are satisfied at x. Criterion (5.8) is a commonly used one which means no significant progress can be made. We introduced $\alpha^k \geq 0.1$ since we did not want to stop the iteration if (5.8) was caused by a very small step length. In our experiments, $tol = 10^{-12}$. This stopping criteria was successful for our experiments since for every problem we tested, the first-order and second-order optimal conditions were confirmed.

Computation and Parameter Setting: Our implementation is straightforward: i.e., w^k is computed by (2.33), p_g^k is computed by (2.35), and p^k is computed by (2.13), etc..

We compute $s^k = \sigma^k \alpha^k p^k + (1-\sigma^k)\alpha_g^k p_g^k$, where σ^k is given by (3.75) and we set $\tau = 0.95$.

Here are the settings we used in our experiments:

$$\tau_1 = 0.8; \ \tau_2 = 0.9999; \ \tau_3 = 0.001; \ \tau_4 = \tau_5 = 0.5. \tag{5.11}$$

We adjust Δ^k as follows:

Updating Δ^k

Let $0 < \tau_6 < \tau_7 < 1$, $0 < \delta_1 < 1 < \delta_2$, and $0 < \Delta_l < \Delta_u$ be given,
$\Delta^0 = 1$;
For $k = 0, 1, 2, \cdots$

$bd^k = \sigma^k \alpha^k + (1-\sigma^k)\alpha_g^k$;

If $bd^k \leq \tau_6$
$\Delta^{k+1} = \max(\Delta_l, \delta_1 \Delta^k)$;
end.
If $bd^k \geq \tau_7$
$\Delta^{k+1} = \min(\Delta_u, \delta_2 \Delta^k)$;
end.

The corresponding parameters are set as follows:

$$\tau_6 = 0.5; \ \tau_7 = 0.9;$$
$$\delta_1 = 0.75; \ \delta_2 = 1.25;$$
$$\Delta_l = 0.1; \ \Delta_u = 10;$$

Experimental Results: We tested four groups of problems using Algorithm **Interior-Newton$_2$** and the results are tabulated in Tables 1 – 4. For each case, three problems were attempted. The table entries are the numbers of iterations required to satisfy the stopping criteria. Positive definite problems are reported in Tables 1 and 2; indefinite problems are reported in Tables 3 and 4. For each problem in Tables 3 and 4, about 10% of the eigenvalues of H are negative.

The lower bounds were all set to zero. The upper bounds were all set to unity in Table 1 and 3 while in Table 2 and 4, about 10% were set to be infinity (that is what

$pctinf$ = 0.1 means). For all the problems, about $0.8 * (n - m)$ components of the generated x were set to the bounds and $cond(H) = cond(A) = cond$.

We did not encounter any instance where $q(x^k) \longrightarrow -\infty$; however, for our generated indefinite problems, with some infinite bounds, there is no guarantee that q is bounded below in the feasible region.

Very few iterations needed more than a single QR factorization (see Theorem 8.10). Over all 3806 iterations (for the 216 test problems), only 18 extra QR factorizations were used. On average, $D^k = (\tilde{V}^k)^{\frac{1}{2}}$ was used about twice per problem (remember that D^1 is always set to be $(\tilde{V}^1)^{\frac{1}{2}}$). The average cost for finding a feasible starting point was 4.30 linear system solves, each of order $(m + n)$.

TABLE 1

Positive Definite Problems, pctinf = 0

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
100	10	18	17	18	16.3	20	17
100	50	16	13	17	15.3	23	17.3
100	90	12	10.3	13	11.3	18	16.7

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
200	20	17	16	21	18.7	23	21
200	100	20	18.3	22	19.3	22	19.3
200	180	15	13.3	16	13.3	17	15.3

TABLE 2

Positive Definite Problems, pctinf = 0.1

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
100	10	18	16.3	17	16	18	16.3
100	50	15	12.7	18	16	22	16
100	90	13	11	14	11.7	16	14

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
200	20	17	16	22	20	19	18.3
200	100	21	18.3	21	19.7	22	19
200	180	15	13.7	14	13	24	17.3

TABLE 3

Indefinite Problems, pctinf = 0

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
100	10	20	18.3	26	20.3	29	23.3
100	50	20	19.3	23	20	17	16
100	90	13	11.3	13	11.7	18	14.7

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
200	20	21	20.3	26	24.7	32	28.7
200	100	25	21.7	28	23.7	23	21.3
200	180	17	15.7	14	12.7	17	16

TABLE 4

Indefinite Problems, pctinf = 0.1

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
100	10	19	17.7	26	20.3	34	24.3
100	50	20	18.3	24	21.3	19	17.3
100	90	13	11.3	14	12.3	17	14.3

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
200	20	39	26	28	25.3	30	30
200	100	27	22	29	23.3	25	21.7
200	180	16	15.3	19	15.7	25	18

Observations: On the whole, the experiments indicate that algorithm **Interior-Newton$_2$** is efficient. The iteration numbers are insensitive to problem size and condition number. Though our experiments are limited, they clearly indicate that the new algorithm is promising.

For purposes of comparison, we have used Algorithm **Interior-Newton$_1$** to solve the same set of problems as in Table 1, and the results are presented in the following table. In most cases both algorithms terminated at (almost) the same function values for each problem; however, the required numbers of iterations are quite different. On average, the number of iterations is 16.1 in Table 1 but is 31.0 in Table 1′. The largest number is 23 in Table 1 but is 89 in Table 1′.

TABLE 1′

Positive Definite Problems, pctinf = 0

($alt^k = 1$ for all k)

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
100	10	20	18	27	20.7	35	27
100	50	60	31.3	69	36.7	28	19
100	90	53	34.3	21	15.7	89	58

size		cond=10^3		cond=10^6		cond=10^9	
n	m	max	avg	max	avg	max	avg
200	20	20	19	43	27.3	40	31
200	100	59	50.3	56	35.7	59	35
200	180	53	44	56	28.7	40	25.7

6. Concluding Remarks. We have proposed an interior Newton method with a new scaling strategy for general quadratic programming problems. The algorithm is robust and has stronger convergence properties than existing interior methods for the general quadratic programming problem. Specifically, the main theoretical property of our proposed method can be summarized as follows. Under compactness and nondegeneracy assumptions, i.e., **(AS1)**, **(AS2)**, and **(AS3)**, our proposed algorithm generates a sequence $\{x^k\}$ converging to a point x^* satisfying first-order and second-order necessary conditions. Moreover, if x^* satisfies second-order sufficiency conditions, then the local rate of convergence is 2-step quadratic. Nevertheless, it may be possible to weaken some of the nondegeneracy assumptions, especially for the first-order convergence results. This is a subject of future research.

Preliminary numerical experiments suggest that the method is efficient for dense problems of moderate size. Inspection of the conditioning of the underlying linear systems, in the limit, indicates that robust asymptotic behavior is to be expected from this approach, even in the presence of near-degeneracy and ill-conditioning. This is in sharp contrast to the limiting linear system in Ye's approach [24] which always approaches singularity. To further explore this remark, consider the following.

Suppose $H > 0$ in $\mathcal{N}^*$. Then for all k sufficiently large, the resulting coefficient matrix of Ye's affine scaling method can be formulated as

$$(\hat{Z}^k)^T \tilde{V}^k H \tilde{V}^k \hat{Z}^k, \tag{6.1}$$

where

$$< \hat{Z}^k >= null(A\tilde{V}^k), \text{ and the columns of } \hat{Z}^k \text{ are orthonormal.} \tag{6.2}$$

Alternatively, the underlying system can be formulated with the following coefficient matrix:

$$\Sigma_1^k \stackrel{\text{def}}{=} \begin{bmatrix} \tilde{V}^k H \tilde{V}^k & (A\tilde{V}^k)^T \\ A\tilde{V}^k & 0 \end{bmatrix}. \tag{6.3}$$

The limit matrix of (6.3),

$$\Sigma_1^* = \lim \Sigma_1^k = \begin{bmatrix} \tilde{V}^* H \tilde{V}^* & (A\tilde{V}^*)^T \\ A\tilde{V}^* & 0 \end{bmatrix} \in \mathcal{R}^{(m+n)\times(m+n)}, \tag{6.4}$$

is singular: it is easy to see that there exists a permutation matrix P_1 such that

$$P_1^T \Sigma_1^* P_1 = \begin{bmatrix} 0 & 0 \\ 0 & B_1 \end{bmatrix} \tag{6.5}$$

for some $B_1 \in \mathcal{R}^{(m+n-|\mathcal{A}^*|)\times(m+n-|\mathcal{A}^*|)}$. Therefore, the limit matrix

$$(\hat{Z}^*)^T \tilde{V}^* H \tilde{V}^* \hat{Z}^* \tag{6.6}$$

of (6.1) is singular. When $\tilde{v}_i^k \longrightarrow 0$, the *ith* row and *ith* column of Σ_1^k go to zero at the rate of $O(\tilde{v}_i^k)$. This ill-conditioning of Σ_1^k is unavoidable no matter how well H and A are conditioned.

In contrast, for all k sufficiently large, the resulting coefficient matrix of our algorithm can be formulated as

$$(\bar{Z}^k)^T \bar{M}^k \bar{Z}^k = (\bar{Z}^k)^T (D^k H D^k + |G^k|)\bar{Z}^k. \tag{6.7}$$

Alternatively, the underlying system can be solved using the coefficient matrix

$$\Sigma_2^k \stackrel{\text{def}}{=} \begin{bmatrix} D^k H D^k + |G^k| & (AD^k)^T \\ AD^k & 0 \end{bmatrix}. \tag{6.8}$$

The limit matrix of (6.8),

$$\Sigma_2^* = \lim \Sigma_2^k = \begin{bmatrix} D^* H D^* + |G^*| & (AD^*)^T \\ AD^* & 0 \end{bmatrix} \in \mathcal{R}^{(m+n)\times(m+n)}, \tag{6.9}$$

is well-behaved under **(AS3)** since there exists a permutation matrix P_2 such that

$$P_2^T \Sigma_2^* P_2 = \begin{bmatrix} diag(|g_{\mathcal{A}^*}^*|) & 0 \\ 0 & B_2 \end{bmatrix} \tag{6.10}$$

for some $B_2 \in \mathcal{R}^{(m+n-|\mathcal{A}^*|)\times(m+n-|\mathcal{A}^*|)}$ and B_2 is nonsingular. The limit matrix

$$(\bar{Z}^*)^T (D^* H D^* + |G^*|)\bar{Z}^* \tag{6.11}$$

of (6.7) is well-conditioned under **(AS3)**. Any possible ill-conditioning of Σ_2^k can only come from the ill-conditioning of H, the rank deficiency of A, and the degeneracy of the original problem (1.1). Moreover, even in the the degenerate case, i.e., for some i, $\tilde{v}_i^* + |g_i^*| = 0$, the *ith* row and *ith* column of Σ_2^k go to zero at the rate of $O((\tilde{v}_i^k)^{\frac{1}{2}})$, much slower than $O(\tilde{v}_i^k)$.

The interior approach we have explored in this paper is an exciting new way to solve quadratic programming problems: it is theoretically interesting and practically viable. As it stands it is not suitable for large-scale quadratic programming in that it requires full-dimensional trust region computations and several matrix factorizations in each major iteration. In response to this we are currently investigating a modification of this approach involving iterative and approximate linear solvers. This will be the topic of a future report.

7. Acknowledgements. We thank our colleague Yuying Li for many helpful remarks on this work. We also thank Yinyu Ye for reading a preliminary draft of this report and making a number of useful suggestions.

8. Appendix: The convergence properties of Algorithm Interior-Newton$_2$. In this appendix, we verify that algorithm **Interior-Newton$_2$** has the same convergence properties as algorithm **Interior-Newton$_1$**. To be specific, and simplify the presentation, we will assume that w^k is computed by (2.33) and p_g^k is computed by (2.35). We are content with this assumption because it can be satisfied easily, i.e., a single least-squares solve in each iteration. On the other hand, it appears the convergence results in this appendix remain valid if this assumption is replaced by more general conditions, e.g., **(CD1)**, **(CD2)**, etc..

The first result we establish is that $\{q(x^k)\}$ converges if w^k, p_g^k are computed by (2.33), (2.35) and s^k satisfies (5.7) for all k sufficiently large. However, Theorem 3.2 may not be true since D^k is not always equal to $(\tilde{V}^k)^{\frac{1}{2}}$. Instead, we will show directly that (3.21) is true for any limit point of $\{x^k\}$, and that (3.55) holds. These two results imply the convergence of $\{x^k\}$. First, we give a simple general result; the proof is straightforward and so it is omitted.

LEMMA 8.1. *Suppose $H \geq 0$ is any symmetric matrix. Then*

$$\|Hx\|^2 \leq x^T H x \, \|H\|. \tag{8.1}$$

LEMMA 8.2. *Suppose w^k is computed by (2.33). Then $\{\|w^k\|\}$ and $\{\lambda_g^k\}$ are bounded, where λ_g^k is defined by (2.38).*

Proof. The boundedness of $\{\|w^k\|\}$ can be seen from (3.5), (3.6) and the way w^k is computed. To consider $\{\lambda_g^k\}$, note that by (2.38), we have

$$\lambda_g^k \mu^k = -(y^k)^T \bar{M}^k y^k \mu^k - (y^k)^T \bar{g}^k, \ \forall k.$$

So, by the boundedness of $\{\|w^k\|\}$ and **(AS1)**, $\exists C_4 > 0$ such that

$$\lambda_g^k |\mu^k| \leq C_4. \tag{8.2}$$

Hence, by (2.40), λ_g^k is bounded. ∎

The following set will be useful subsequently:

$$\mathcal{K} \stackrel{\text{def}}{=} \{k : D^k = (\tilde{V}^k)^{\frac{1}{2}} \text{ at the end of Step 1}\}. \tag{8.3}$$

We let $\mathcal{K}^c$ denote the complementary set of $\mathcal{K}$. It is clear that $k \in \mathcal{K}^c$ if and only if $D^k = |V^k|^{\frac{1}{2}}$ at the end of Step 1.

THEOREM 8.3. *Suppose w^k and p_g^k are computed by (2.33) and (2.35) for all k sufficiently large. Suppose $\{k_j\} \subset \mathcal{K}^c$ and $x^{k_j} \longrightarrow x^*$. Then (3.21) holds and for all k_j sufficiently large,*

$$(8.4) \qquad \check{v}_i^{k_j} = \tilde{v}_i^{k_j}, \ \forall i \in \mathcal{A}^*,$$

where $\check{v}^k$ is defined by (5.4).

Proof. First, assume that zero is a limit point of $\{\check{t}^{k_j}\}$, where $\check{t}^k$ is defined by (5.6). Without loss of generality, assume $\check{t}^{k_j} \longrightarrow 0$. Then

$$(8.5) \qquad \check{v}^{k_j}.*\check{g}^{k_j} \longrightarrow 0.$$

By the boundedness of $\{\|w^k\|\}$ and (1.14), both $\{w^{k_j-1}\}$ and $\{\check{v}_i^{k_j}\}$ have at least one limit point. Assume $w^{k_j-1} \longrightarrow w^*$ and $\check{v}_i^{k_j} \longrightarrow \check{v}^*$. Then

$$(8.6) \qquad \check{v}^*.*g^* = 0, \text{ where } g^* \stackrel{\text{def}}{=} Hx^* + C + A^T w^*.$$

So by (1.15), $\tilde{v}^*.*g^* = 0$ which implies that $w^* = w_{ls}^*$ by **(AS2)**. Hence $g^* = g_{ls}^*$ and

$$(8.7) \qquad \tilde{v}^*.*g_{ls}^* = \check{v}^*.*g_{ls}^* = 0.$$

Now assume that $\{\check{t}^{k_j}\}$ is bounded away from zero. Then for some $\epsilon > 0$,

$$(8.8) \qquad \check{t}^{k_j} \ge \epsilon, \ \forall k_j.$$

Since $k_j \in \mathcal{K}^c$, by the convergence of $\{q(x^k)\}$, (3.2), (5.7), and by the acceptance test in Step 1, we have

$$(8.9) \qquad \begin{aligned} q(x^{k_j}) - q(x^{k_j+1}) &\ge -\gamma_2 \psi^{k_j}(\alpha_g^{k_j} p_g^{k_j}) \\ &\ge \gamma_2 \min(\tau_3, \tau_4\epsilon)(-\psi^{k_j}(p_g^{k_j})) \longrightarrow 0. \end{aligned}$$

So, by (3.3),

$$(8.10) \qquad ((y^{k_j})^T \bar{M}^{k_j} y^{k_j} + \lambda_g^{k_j})(\mu_g^{k_j})^2 \longrightarrow 0.$$

Considering Lemma 8.2, we see that $\|(y^{k_j})^T \bar{M}^{k_j} y^{k_j} + \lambda_g^{k_j}\|$ is bounded. So (2.38) and Lemma 8.1 yield that (remember that $AD^{k_j}\bar{Z}^{k_j} = 0$)

$$(8.11) \qquad \begin{aligned} &\|(\bar{Z}^{k_j})^T D^{k_j}(Hx^{k_j} + c)\| = \|(\bar{Z}^{k_j})^T \bar{g}^{k_j}\| \\ = \ &(y^{k_j})^T \bar{g}^{k_j} = -((y^{k_j})^T \bar{M}^{k_j} y^{k_j} + \lambda_g^{k_j})\mu_g^{k_j} \longrightarrow 0, \end{aligned}$$

where $D^{k_j} = |\check{V}^{k_j}|^{\frac{1}{2}}$. Let $\check{v}^*$ be any limit point of $\{\check{v}^{k_j}\}$ and assume that $\check{v}^{k_{j_l}} \longrightarrow \check{v}^*$. Let $\check{\bar{Z}}^* = \lim \bar{Z}^{k_{j_l}}$. Then

$$(8.12) \qquad (\check{\bar{Z}}^*)^T \check{D}^*(Hx^* + c) = 0.$$

Since $< \check{\bar{Z}}^* >= null(A\check{D}^*)$, $\exists w^* \in \mathcal{R}^m$ such that

$$\check{D}^*(Hx^* + c) + (A\check{D}^*)^T w^* = 0,$$

i.e.,

$$D̆^* g^* = \check{D}^*(Hx^* + c + A^T w^*) = 0, \tag{8.13}$$

which gives us (8.6) again and consequently, (8.7) is true.

To show (8.4), let $\check{v}^*$ be any limit point of $\{\check{v}^{k_j}\}$. Repeat the arguments above, we have

$$\check{v}^* .* g_{ls}^* = 0. \tag{8.14}$$

Then, $\forall i \in \mathcal{A}^*$, $\check{v}_i^* = \tilde{v}_i^* = 0$. Therefore, for all k_j sufficiently large, (8.4) holds. ∎

THEOREM 8.4. *Suppose w^k and p_g^k are computed by (2.33) and (2.35) for all k sufficiently large. Suppose x^* is any limit point of $\{x^k\}$ and $x^{k_j} \longrightarrow x^*$. Then (3.21) holds. Furthermore, if s^k satisfies* **(CD3)** *with $D^k = (\tilde{V}^k)^{\frac{1}{2}}$ or $D^k = |\check{V}^k|^{\frac{1}{2}}$ for all k sufficiently large, then $\{x^k\}$ converges to x^* which satisfies (3.48).*

Proof. If $k_j \in \mathcal{K}$, then s^{k_j} is computed based on $D^{k_j} = (\tilde{V}^{k_j})^{\frac{1}{2}}$, and otherwise, based on $D^{k_j} = |\check{V}^{k_j}|^{\frac{1}{2}}$. So if there are infinitely many $k_j \in \mathcal{K}^c$, then by Theorem 8.3, (3.21) holds. Otherwise, by (5.7) and Corollary 3.3, (3.21) holds also.

To see the convergence of $\{x^k\}$, we first show that (3.55) is true. In fact, since s^k satisfies **(CD3)** with $D^k = (\tilde{V}^k)^{\frac{1}{2}}$ or $D^k = |\check{V}^k|^{\frac{1}{2}}$ for all k sufficiently large, (3.55) is obviously true for those $k \in \mathcal{K}$ and is also true for those $k \in \mathcal{K}^c$ by (8.4). Note that the conditions used by Lemma 3.4, Lemma 3.5 and Theorem 3.6 are only (3.21) and (3.55), we see that $\{x^k\}$ converges to x^* which satisfies (3.48). ∎

Note that the convergence of $\{x^k\}$ does not depend on whether $D^k = (\tilde{V}^k)^{\frac{1}{2}}$ or $D^k = |\check{V}^k|^{\frac{1}{2}}$ for the trial D^k in Step 1. This flexibility gives us freedom to choose D^k with performance in mind.

The next results establish the convergence of $\{w^k\}$, and the convergence of $\{x^k\}$ to a first-order point.

COROLLARY 8.5. *The sequence $\{w^k\}$ converges to w_{ls}^* under the assumptions of Theorem 8.4.*

Proof. Let w^* be any limit point of $\{w^k\}$ and assume that $\{w^{k_j}\} \longrightarrow w^*$. If there are infinitely many $k_j \in \mathcal{K}$, then $w^* = w_{ls}^*$. Otherwise, for all k_j sufficiently large, $k_j \in \mathcal{K}^c$ and by (2.33),

$$w^{k_j} = -(A|\check{V}^{k_j}|A^T)^{-1} A|\check{V}^{k_j}|(Hx^{k_j} + c). \tag{8.15}$$

Let $\check{v}^*$ be any limit point of $\{\check{v}^{k_j}\}$ and assume that $\check{v}^{k_{j_l}} \longrightarrow \check{v}^*$. Then by (8.14), (3.6) and (8.15), we have

$$\begin{aligned} w_{ls}^* &= -(A|\check{V}^*|A^T)^{-1} A|\check{V}^*|(Hx^* + c) \\ &= \lim w^{k_{j_l}} = w^*. \end{aligned} \tag{8.16}$$

∎

LEMMA 8.6. *If s^k satisfies (3.60) for all $k \in \mathcal{K}$ sufficiently large, the limit point (x^*, w_{ls}^*) satisfies (1.4).*

Proof. If there are infinitely many $k \in \mathcal{K}^c$, say $\{k_j\} \subset \mathcal{K}^c$, then $\check{g}^{k_j} \longrightarrow g_{ls}^*$ by Corollary 8.5. Let $\check{v}^*$ be any limit point of $\{\check{v}^{k_j}\}$. By Theorem 8.3, $\check{v}^*. * g_{ls}^* = 0$ and $\forall i \in \mathcal{A}^*$, $v_i^* = \check{v}_i^* = 0$. So $v^*. * g_{ls}^* = 0$. That is (1.4).

If $k \in \mathcal{K}$ for all k sufficiently large, then we are identically using Algorithm **Interior-Newton$_1$**, and, by Corollary 3.7, the result holds as well. ∎

Comparing Lemma 8.6 to Lemma 3.7, we see that Corollary 8.6 is stronger since (3.60) needs to be satisfied only by those s^k with $k \in \mathcal{K}$ to get a first-order limit point. As indicated in Lemma 3.8, one possible strategy to guarantee first-order convergence is to replace the direction s^k with the scaled gradient p_g^k whenever (3.60) fails. Alternatively, satisfaction of **(CD4)** will yield first-order convergence, and more.

The following result, with proof similar to the proof of Theorem 3.12, establishes convergence to a second-order point.

THEOREM 8.7. *Suppose w^k is computed by (2.33) and $\alpha^k p^k$ satisfies* **(CD4)** *for all k sufficiently large. Further, assume $\{x^k\}$ converges to a point x^* that satisfies (3.21). Then x^* satisfies (1.5).*

It is easy to verify that all the remaining results in Sections 3 and 4 for **Interior-Newton$_1$** also hold for **Interior-Newton$_2$**. To sum them up, we have the following results:

THEOREM 8.8. *Suppose for all k sufficiently large, w^k and p_g^k are computed by (2.33) and (2.35), s^k satisfies (3.60) and* **(CD3)** *with $D^k = (\tilde{V}^k)^{\frac{1}{2}}$ or $D^k = |\check{V}^k|^{\frac{1}{2}}$, and $\alpha^k p^k$ satisfies* **(CD4)**. *Then the sequence $\{x^k\}$ generated by Algorithm* **Interior-Newton$_2$** *converges to a point x^* which satisfies the first-order and second-order necessary optimality conditions. Further more, if $H > 0$ in $\mathcal{N}^*$, then x^* is a solution to (1.1).*

The next result establishes that the point of convergence is a second-order point; moreover, the rate of convergence is at least 2-step quadratic. (Recall that (4.47) implies a 2-step quadratic convergence rate.)

COROLLARY 8.9. *Suppose for all k sufficiently large, w^k and p_g^k are computed by (2.33) and (2.35), $s^k = \sigma^k \alpha^k p^k + (1 - \sigma^k)\alpha_g^k p_g^k$, where σ^k is defined by (3.75) and α^k, p^k and α_g^k are defined by (2.27), (2.13) and (2.41). Then the sequence $\{x^k\}$ generated by Algorithm* **Interior-Newton$_2$** *converges to a point x^* which satisfies the first-order and second-order necessary optimality conditions. Further more, if $H > 0$ in $\mathcal{N}^*$, then $\{x^k\}$ satisfies (4.47) and x^* is a solution to (1.1).*

Our final result in this appendix is to show that, ultimately, only one computation is needed for w^k, $\bar{Z}^k$, and p_g^k in each iteration. Clearly, this is true if we can establish:

(8.17) The trial D^k is unchanged in Step 1 for all k sufficiently large.

THEOREM 8.10. *Suppose w^k and p_g^k are computed by (2.33) and (2.35). Suppose $x^k \longrightarrow x^*$ and $H > 0$ in $\mathcal{N}^*$. Then (8.17) holds.*

Proof. First, we show the following:

$$\exists \alpha_0 > 0 \text{ such that } \alpha_g^k \geq \alpha_0, \forall k. \tag{8.18}$$

In fact, by (8.4), whether $D^k = (\tilde{V}^k)^{\frac{1}{2}}$ or $D^k = |\check{V}^k|^{\frac{1}{2}}$, we always have that

$$D_{ii}^k = (\tilde{v}_i^k)^{\frac{1}{2}}, \ \forall i \in \mathcal{A}^*, \ \forall k \text{ sufficiently large.}$$

If (8.18) is false, then repeat the arguments between (3.19) and (3.20), note that $\|g_{ls}^{k_j}\|$ is bounded, we would have

$$\frac{\|D^{k_j} g_{ls}^{k_j}\|}{|\mu_g^{k_j}|} \longrightarrow 0. \tag{8.19}$$

But on the other hand, since $AD^2 g_{ls} = 0$, $\exists \xi \in \mathcal{R}^{n-m}$ such that $Dg_{ls} = \bar{Z}\xi$. Hence

$$y^T \bar{M} y = \frac{1}{\|Dg_{ls}\|^2} g_{ls}^T D^2 M D^2 g_{ls} = \frac{\xi^T \bar{Z}^T \bar{M} \bar{Z} \xi}{\|\xi\|^2}, \tag{8.20}$$

where y is defined by (2.34). By Lemma 3.9, $(\bar{Z}^*)^T \bar{M}^* \bar{Z}^* > 0$. So $\exists \epsilon > 0$ such that

$$(y^k)^T \bar{M}^k y^k \geq \epsilon. \tag{8.21}$$

Then, using (3.17), we have

$$\frac{\|D^{k_j} g_{ls}^{k_j}\|}{|\mu_g^{k_j}|} \geq \epsilon,$$

which is a contradiction to (8.19). Hence, (8.18) is true.

Now, to show (8.17), it suffices to show that $\{\psi^k(\alpha_g^k p_g^k)/\psi^k(p_g^k)\}$ is bounded away from zero since $\check{t}^k \longrightarrow 0$. This is true because it is easy to see that $(p_g^k)^T g^k \leq 0$, so $\alpha_g^k (p_g^k)^T g^k \leq (\alpha_g^k)^2 (p_g^k)^T g^k$ since $\alpha_g^k \leq 1$, and then, by (8.18),

$$\frac{\psi^k(\alpha_g^k p_g^k)}{\psi^k(p_g^k)} \geq \frac{(\alpha_g^k)^2 \, \psi^k(p_g^k)}{\psi^k(p_g^k)} \geq \alpha_0^2 \tag{8.22}$$

for all k sufficiently large. ∎

REFERENCES

[1] E. M. L. Beale, *On quadratic programming*, Naval Research Logistics Quarterly, 6 (1959), pp. 227-243.

[2] P.H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, Mathematical Programming, 39 (1987), pp. 93-116.

[3] T. F. Coleman and L. A. Hulbert, *A direct active set algorithm for large sparse quadratic programs with simple bounds*, Mathematical Programming, 45 (1989), pp. 373-406.

[4] T. F. Coleman and Y. Li, *On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds*, Tech. Rep. TR 92-1314, Computer Science Department, Cornell University, 1992.

[5] T. F. Coleman and Y. Li, *An interior trust region approach for nonlinear minimization subject to bounds*, Tech. Rep. TR 93-1342, Computer Science Department, Cornell University, 1993

[6] R. W. Cottle and G. B. Dantzig, *Complementary pivot theory of mathematical programming*, Linear Algebra Appl., 1 (1968), pp. 103-125.

[7] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[8] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk SSSR*, 174:747–748, 1967. Translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967.

[9] R. Fletcher, *A general quadratic programming algorithm*, Journal of the Institute of Mathematics and Its Applications, 7 (1971), pp. 76-91.

[10] P. E. Gill and W. Murray, *Newton type methods for unconstrained and linearly constrained optimization*, Mathematical Programming, 7 (1974), pp. 311-350.

[11] D. Goldfarb, *Extensions of Newton's method and simplex methods for solving quadratic programs*, in Numerical Methods for nonlinear Optimization, F. A. Lootsman, eds., Academic Press, London, 1972.

[12] N. Karmarkar, *A new polynomial time algorithm for linear programming*, Combinatorica, 4 (1984) pp. 373-395.

[13] J. Liu, *Interior and Exterior Newton Methods for Quadratic Programming*, Ph.D. Thesis, Center for Applied Mathematics, Cornell University, Ithaca, NY, 1994.

[14] J. J. Moré and D. Sorensen, *Computing a trust region step*, SIAM Journal on Scientific and Statistical Computing, 4 (1983), pp. 553-572.

[15] J. J. Moré and G. Toraldo, *Algorithms for bound constrained quadratic programming problems*, Numerische Mathematik, 55 (1989), pp. 377-400.

[16] K. G. Murty and S. N. Kabadi, *Some NP-complete problems in quadratic and nonlinear programming*, Mathematical Programming, 19:200-212, 1987.

[17] P. M. Pardalos, Y. Ye, and C. G. Han. An interior point algorithm for large–scale quadratic problems with box constraints. In A. Bensoussan and J. L. Lions, editors, *Analysis and Optimization of Systems : Proceedings of the 9th International Conference, Antibes, France, 1990*, volume 144 of *Lecture Notes in in Control and Information Science*, pages 413–422. Springer Verlag, Berlin, West–Germany, 1990.

[18] M. J. D. Powell, *A new algorithm for unconstrained optimization*, in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter, eds., Academic Press, New York, 1970, pp. 31-65.

[19] G. A. Schultz, R. B. Schnabel, and R. H. Byrd, *A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties*, SIAM Journal on Numerical Analysis, 22 (1985), pp. 47-67.

[20] D. Sorensen, *Trust region methods for unconstrained optimization*, SIAM Journal on Numerical Analysis, 19 (1982), pp. 409-426.

[21] S.A. Vavasis, *Nonlinear Optimization: Complexity Issues*, (Oxford University Press, 1991)

[22] S. J. Wright, *An interior-point algorithm for linearly constrained optimization*, SIAM Journal on Optimization, 2 (1992), pp. 450-473.

[23] Y. Ye, *An extension of Karmarkar's algorithm and the trust region method for quadratic programming*, in: N. Megiddo, ed., *Progress in Mathematical Programming* (Springer, New York, 1989).

[24] Y. Ye, *On affine scaling algorithms for nonconvex quadratic programming*, Mathematical Programming, 56 (1992) 285-300.