On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds [1]

Thomas F. Coleman and Yuying Li
Computer Science Department
and
Advanced Computing Research Institute[2]
Cornell University
Ithaca, New York 14853

December 7, 1992

Abstract. We consider a new algorithm, a reflective Newton method, for the problem of minimizing a smooth nonlinear function of many variables, subject to upper and/or lower bounds on some of the variables. This approach generates strictly feasible iterates by following piecewise linear paths ("reflection" paths) to generate improved iterates. The reflective Newton approach does not require identification of an "activity set". In this report we establish that the reflective Newton approach is globally and quadratically convergent. Moreover, we develop a specific example of this general reflective path approach suitable for large-scale and sparse problems.

[1] 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 in part by NSF, AFOSR, and ONR through grant DMS-8920550, and by 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.

[2] The Advanced Computing Research Institute is a unit of the Cornell Center for Theory and Simulation in Science and Engineering (Theory Center).

1. Introduction. This paper is concerned with minimizing a smooth nonlinear function subject to bounds on the variables:

$$\min_{x \in \Re^n} f(x), \quad l \leq x \leq u, \tag{1.1}$$

where $l \in \{\Re \cup \{-\infty\}\}^n$, $u \in \{\Re \cup \{\infty\}\}^n$, $l \leq u$, and $f : \Re^n \rightarrow \Re^1$. We denote the feasible set $\mathcal{F} = \{x : \; l \leq x \leq u\}$ and the strict interior $int(\mathcal{F}) = \{x : \; l < x < u\}$.

Minimization problems with upper and/or bounds on some of the variables form an important and common class of problems. There are many algorithms for this type of optimization problem, some of which are restricted to quadratic (in some cases convex quadratic) objective functions and some are more general (e.g., [2, 5, 11, 13, 14, 15, 20, 22, 23, 24, 25, 29]). However, in contrast to the new approach we analyze here, few of these approaches represent efficient ways to solve large-scale nonlinear problems to high accuracy.

The main purpose of this paper is to consider the convergence properties of a new reflective Newton approach, introduced in [10] for the case where f is a quadratic function. In particular, here we establish that reflective Newton methods, applied to twice continuously-differentiable nonlinear functions f, are globally and quadratically convergent under reasonable assumptions.

Reflective Newton methods appear to have significant practical potential for large-scale problems. Consider, for example, the results quoted [10] for the "obstacle problem" on a square m-by-m mesh – see Table 1. The column "its" refers to the number of iterations required to achieve an accurate solution – the cost of each iteration is roughly proportional to the cost of a sparse Cholesky factorization of an n-by-n sparse symmetric positive definite matrix. Full details are given in [10].

TABLE 1
Obstacle Problem: Lower and Upper Bounds

m	n	its
30	900	11
40	1600	12
50	2500	14
60	3600	13
100	10,000	14

A remarkable feature of this type of algorithm, illustrated by this typical example, is the very slow growth in required number of iterations. Given a class of problems and a "natural" way to increase the problem dimension, reflective Newton methods appear to be strikingly insensitive to problem size. Experiments reported in [10] are restricted to quadratic problems; we are currently experimenting on more general nonlinear problems and preliminary results continue to support this claim.

A *reflective algorithm* for problem (1.1) is an algorithm that uses the *reflective transformation* to maintain feasibility [10]. For a problem with nonnegativity constraints only, $\mathcal{F} = \{x : x \geq 0\}$, a reflective mapping is merely the absolute value function,

$R : \mathcal{R}^n \overset{onto}{\rightarrow} \mathcal{F}$, i.e., $x = R(y) = |y|$, where the absolute value notation is meant to apply to each component. More generally, a reflective mapping (or transformation) for problem (1.1) is an open mapping $R : \mathcal{R}^n \overset{onto}{\rightarrow} \mathcal{F}$ defined in Figure 1. An illustration of a 1-dimensional reflective transformation is given in Figure 2.

Case 1: $(l_i > -\infty, \ u_i < \infty)$

To evaluate $x_i = R(y)_i$:

$w_i = |y_i - l_i| \ \mathbf{mod} \ [2(u_i - l_i)], \ x_i = \min(w_i, 2(u_i - l_i) - w_i) + l_i$

Case 2: $(l_i > -\infty, \ u_i = \infty)$

To evaluate $x_i = R(y)_i$: If $y_i \geq l_i, \ x_i = y_i$, else $x_i = 2l_i - y_i$.

Case 3: $(l_i = -\infty, \ u_i < \infty)$

To evaluate $x_i = R(y)_i$: If $y_i \leq u_i, \ x_i = y_i$, else $x_i = 2u_i - y_i$.

Case 4: $(l_i = -\infty, \ u_i = \infty)$.

In this case there are no constraints on x_i and so $x_i = y_i$.

FIG. 1. *The Reflective Transformation R*

Using this reflective transformation $R(y)$, (1.1) can be replaced with the unconstrained piecewise differentiable problem:

$$\min_{y \in R^n} \hat{f}(y) \tag{1.2}$$

where $\hat{f}(y) = f(R(y))$. A reflective algorithm for the original problem (1.1) is a descent direction algorithm[3] for $\hat{f}(y)$ – see Figure 3. Algorithm 1 generates the sequence $\{y_k\}$; the strictly feasible sequence $\{x_k\}$ can be obtained from the relation $x_k = R(y_k)$. (Note: strict feasibility is maintained because the line search does not accept breakpoints – breakpoints correspond to points on the boundary.)

The straight-line direction s_k^y corresponds to a piecewise linear path in x-space. This piecewise linear path can be described, recursively, as follows.

For simplicity, and without loss of generality, assume $y_k = x_k$. Define the vector[4]

$$BR_k = \max[(l - x_k) \ ./ \ s_k^y, \quad (u - x_k) \ ./ \ s_k^y)], \tag{1.3}$$

where the notation " ./ " indicates componentwise division. Component i of vector BR_k records the positive distance form x_k to the breakpoint corresponding to variable

---

[3] Direction s_k^y is a descent direction for for $\hat{f}(y)$ at y_k if $\hat{f}(y_k + \alpha s_k^y) < \hat{f}(y_k)$ for all positive sufficiently small α.

[4] For the purpose of computing BR we assume the following rules regarding arithmetic with infinities. If a is a finite scalar then $a + \infty = \infty$, $a - \infty = -\infty$, $\frac{\infty}{a} = \infty \cdot \mathrm{sgn}(a)$, $\frac{-\infty}{a} = -\infty \cdot \mathrm{sgn}(a)$, $\frac{a}{0} = \mathrm{sgn}(a) \cdot \infty$, $\frac{\infty}{0} = \infty$, *and* $\frac{-\infty}{0} = -\infty$, where $\mathrm{sgn}(a) = +1$ if $a \geq 0$, $\mathrm{sgn}(a) < 0$ if $a < 0$.

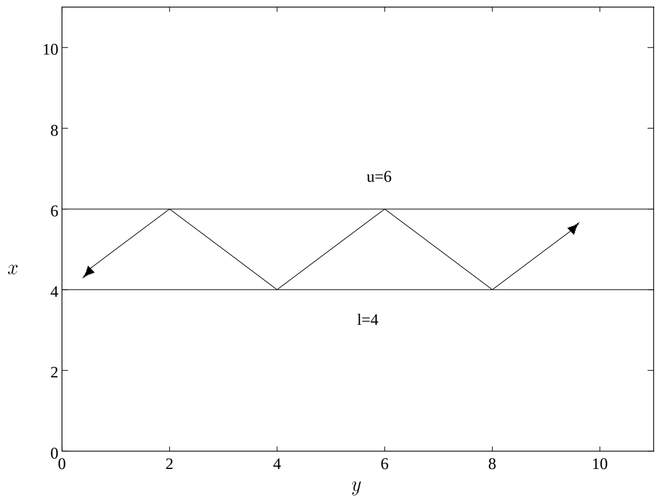

FIG. 2. *A 1-Dimensional Reflective Transformation Example*

Algorithm 1

Choose $y_1 \in int(\mathcal{F})$.
For $k = 1, 2, ...$

1. Determine a descent direction s_k^y for $\hat{f}(y)$ at y_k
2. Perform an approximate line minimization of $\hat{f}(y_k + \alpha s_k^y)$, with respect to α, to determine an acceptable stepsize α_k (such that α_k does not correspond to a breakpoint)
3. $y_{k+1} = y_k + \alpha_k s_k^y$

FIG. 3. *Descent dir'n algorithm for* $\hat{f}(y)$

x_{k_i} in the direction s_k^y. The piecewise linear (reflective) path is defined by Algorithm 2. Since only a single outer iteration is considered, we do not include the subscript k with the variables in our description of Algorithm 2 - dependence on k is assumed.

Given the current point x_k and a descent direction s_k^x let $p_k(\alpha)$ denote the piecewise linear path defined by Algorithm 2: For $\beta_k^{i-1} \leq \alpha < \beta_k^i$,

$$p_k(\alpha) = b_k^{i-1} + (\alpha - \beta_k^{i-1}) p_k^i. \tag{1.4}$$

A two dimensional reflective path is illustrated in Figure 5.

Note that it is now possible to describe Algorithm 1 entirely in x-space without explicitly introducing either the function $\hat{f}$ or the variables y. We do this in Algorithm 3 (in Figure 6).

The difference between Algorithm 1 and Algorithm 3 is purely notational. The view presented by Algorithm 3 has the advantage that it is in the original space – visualization of the reflective process is natural. The advantage of the first view, Algorithm 1, is that

---

Algorithm 2 [Let $\beta^0 = 0$, $p^1 = s^x$, set $b^0 = x_k$.]
[i_u is a finite upper bound on the number of segments of the path to be determined]

For $i = 1 : i_u$

1. Let β^i be the distance to the nearest breakpoint along p^i:

$$\beta^i = \min\{BR : \; BR > 0\}$$

2. Define i^{th} breakpoint: $b^i = b^{i-1} + (\beta^i - \beta^{i-1})p^i$.
3. Reflect to get new dir'n and update BR:
   (a) $p^{i+1} = p^i$
   (b) For each j such that $(b^i)_j = u_j$ (or $(b^i)_j = l_j$)
   - $BR(j) = BR(j) + |\frac{u_j - l_j}{(s^x)_j}|$.
   - $(p^{i+1})_j = -(p^i)_j$

FIG. 4. *Determine the linear reflective path p*

---

the algorithm is a straight line descent direction algorithm, a familiar structure. It is probably useful for the reader to keep both views in mind. In this paper we will primarily work in the original space (x-space) and Algorithm 3. For simplicity we now drop the superscript x (e.g., s^x becomes s).

What restrictions on s_k are needed to obtain convergence of Algorithm 3? Clearly s_k needs to be a descent direction for f at x_k. However, this is not enough. The reason for this is that we must get sufficient decrease in f along the path $p_k(\alpha)$: For an arbitrary descent direction s_k the first breakpoint may be a very short step from the current point (along s_k) and there is no guarantee of continued descent past this breakpoint – the result may be insufficient decrease in f to yield a convergence result.

We use two properties defined in Section 3, "constraint compatibility" and "consistency", to ensure that sufficient decrease is always achievable. Moreover, to get second-order convergence we require the use of directions with sufficient negative curvature.

What restrictions on s_k guarantee quadratic convergence? It turns out that there is a Newton system lurking behind the scenes, based on optimality conditions. If we can guarantee that unit steps be taken (with respect to this system), and satisfy all other constraints mentioned above, then quadratic convergence will follow. In Section 5 we show that this can be done. Section 6 is concerned with a practical variation of the basic method suitable for large-scale problems; Section 7 contains concluding remarks and a look ahead.

Notation: For brevity we denote $g = g(x) \stackrel{def}{=} \nabla f(x)$; $g_k \stackrel{def}{=} g(x_k)$; $g_* \stackrel{def}{=} g(x_*) = \nabla f(x_*)$, where x_* is a specified (usually optimal) point.

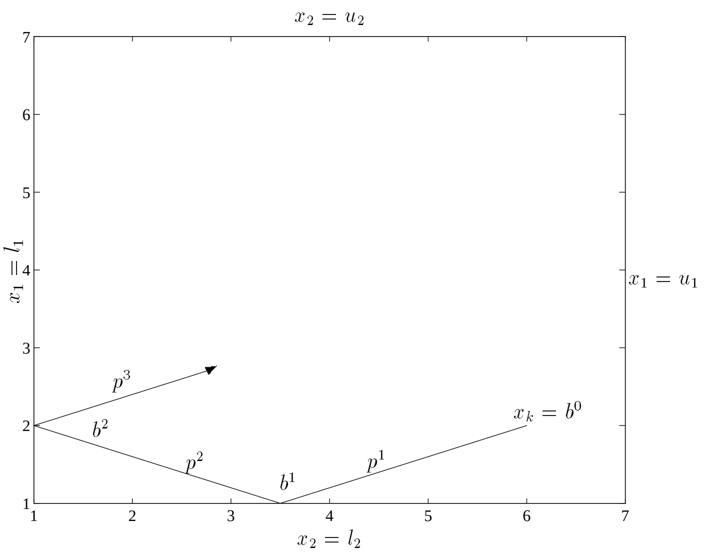

FIG. 5. *A Reflective Path*

Algorithm 3

Choose $x_1 \in int(\mathcal{F})$.

For $k = 1, 2, ...$

1. Determine an initial descent dir'n s_k^x for f at $x_k \in int(\mathcal{F})$. Determine the piecewise linear reflective path $p_k(\alpha)$ via Algorithm 2.
2. Perform an approximate piecewise line minimization of $f(x_k + p_k(\alpha))$, with respect to α, to determine an acceptable stepsize α_k (such that α_k does not correspond to a breakpoint).
3. $x_{k+1} = x_k + p_k(\alpha_k)$.

FIG. 6. *A reflective path algorithm*

Optimality conditions. Optimality conditions for problem (1.1) are well-established. Assuming feasibility, first-order necessary conditions for x_* to be a local minimizer are:

$$\text{first order:} \quad \begin{cases} (g_*)_i = 0 & \text{if } l_i < (x_*)_i < u_i, \\ (g_*)_i \le 0 & \text{if } (x_*)_i = u_i, \\ (g_*)_i \ge 0 & \text{if } (x_*)_i = l_i \end{cases} \tag{1.5}$$

It is interesting to note that the first-order conditions can be expressed as a diagonal system of nonlinear equations, continuous but not everywhere differentiable. To do this we define below a vector $v(x)$ and diagonal matrix $D(x)$, where

$$D^2(x) = \mathrm{diag}(|v(x)|), \tag{1.6}$$

i.e., D^2 is a diagonal matrix with the i^{th} diagonal component equal to $|v_i(x)|$. **The first-order optimality conditions** can be written: If a feasible point x_* is a local

(i) If $g_i < 0$ and $u_i < \infty$ then $v_i = x_i - u_i$.

(ii) If $g_i \geq 0$ and $l_i > -\infty$ then $v_i = x_i - l_i$.

(iii) If $g_i < 0$ and $u_i = \infty$ then $v_i = -1$.

(iv) If $g_i \geq 0$ and $l_i = -\infty$ then $v_i = 1$.

FIG. 7. *Definition of* $v(x)$

minimizer of (1.1) then

$$D_*^2 g_* = 0. \tag{1.7}$$

Second-order conditions involve the Hessian matrix of f, $H = H(x) \stackrel{def}{=} \nabla^2 f(x)$. We assume f is twice continuously-differentiable. Let $Free_*$ denote the set of indices corresponding to "free" variables at point x_*:

$$Free_* = \{i : l_i < (x_*)_i < u_i\}.$$

Second-order necessary conditions can be written[5]: If a feasible point x_* is a local minimizer of (1.1) then $D_*^2 g_* = 0$ and $H_*^{Free_*} \geq 0$ where $H_*^{Free_*}$ is the submatrix of $H_* = H(x_*)$ corresponding to the index set $Free_*$

These conditions are necessary but not sufficient. Sufficiency conditions that are achievable in practise often require a nondegeneracy assumption. This is the case here.

DEFINITION 1. *A point* $x \in \Re^n$ *is* nondegenerate *if, for each index i:*

$$g_i = 0 \implies l_i < x_i < u_i.$$

With this definition we can state **second-order sufficiency conditions**: If a nondegenerate feasible point x_* satisfies $D_*^2 g_* = 0$ and $H_*^{Free_*} > 0$, then x_* is a local minimizer of (1.1).

The theory we develop allows for some latitude in the manner in which a descent direction is obtained. Our particular proposal relies heavily on a (reduced) trust region model to generate directions. In particular, we often determine s_k, at x_k, by solving

$$\min_s \{s^T g_k + \frac{1}{2} s^T M_k s : \; \|D_k^{-1} s\|_2 \leq \Delta_k, \; s \in \mathcal{S}_k\} \tag{1.8}$$

where $\mathcal{S}_k$ is a subspace of $\mathcal{R}^n$, D_k is a positive diagonal scaling matrix, and $\Delta_k > 0$. Appropriate definitions of matrices M and D are crucial to the determination of successful directions. We choose

$$M(x) \;\; = [H + J^v D^{\frac{g}{v}}] \tag{1.9}$$

[5] Notation: If a matrix A is a symmetric matrix then we write $A > 0$ to mean A is positive definite; $A \geq 0$ means A is positive semi-definite.

where H is the Hessian matrix, i.e., $H = H(x) = \nabla^2 f(x)$; J^v is the Jacobian[6] of v, where v is defined in Fig. 5. Matrix $D^{\frac{g}{v}}$ is a diagonal matrix with component i defined $D^{\frac{g}{v}}_{ii} = \frac{g_i^+(x)}{|v_i(x)|}$, for $i = 1 : n$; vector $g^+(x)$ is an "extended gradient", extended to deal with possible degeneracy. In particular,

$$(1.10) \qquad g_i^+ = \begin{cases} |g_i| + \tau_g & \text{if } |g_i| + |v_i|^{\frac{1}{2}} \le \tau_g \\ |g_i| & \text{otherwise} \end{cases}$$

where τ_g is a small positive constant. Clearly if x is a nondegenerate point and τ_g is sufficiently small then $g^+ = |g|$.

The diagonal matrix $D(x)$, used in (1.8), is defined by (1.6), i.e.[7],

$$(1.11) \qquad D(x) = \mathrm{diag}(|v(x)|^{\frac{1}{2}}).$$

Using definition (1.11), problem (1.8) can be written

$$(1.12) \qquad \min_{\bar{s}} \{\bar{s}^T \bar{g}_k + \frac{1}{2}\bar{s}^T \bar{M}_k \bar{s} : \ \|\bar{s}\|_2 \le \Delta_k, \ D_k \bar{s} \in \mathcal{S}_k\}$$

where

$$(1.13) \qquad \bar{M}_k = D_k M_k D_k = D_k H_k D_k + J_k^v D_k^{g^+}, \quad \bar{g}_k = D_k g_k, \quad \bar{s} = D_k^{-1} s,$$

and D^{g^+} is a diagonal matrix, $D^{g^+} = \mathrm{diag}(g^+)$.

Typically subspace $\mathcal{S}_k$ is small, e.g., $|\mathcal{S}_k| = 2$, and the concerns about s_k mentioned above are satisfied by choosing $\mathcal{S}_k$ appropriately. A related reduced trust region idea has been explored in the unconstrained minimization setting [3, 27]. We discuss the definition of $\mathcal{S}_k$ in Section 6. Given $\mathcal{S}_k$, the subspace trust region problem (1.8) or (1.12) can be approached in the following way. Let $\mathcal{S}_k$ be defined by the t_k independent columns of an n-by-t_k matrix V_k, i.e[8], $\mathcal{S}_k = < V_k >$; Therefore, $s = V_k s_v$ for some vector s_v. Let Y_k be an orthonormalization of the columns of $D_k^{-1} V_k$. Hence,

$$D_k^{-1} s = D_k^{-1} V_k s_{V_k} = Y_k s_{Y_k}$$

for some vector s_{Y_k}. Therefore problem (1.8) becomes

$$(1.14) \qquad \min_{s_{Y_k}} \{s_{Y_k}^T Y_k^T \bar{g} + \frac{1}{2} s_{Y_k}^T Y_k^T \bar{M}_k Y_k s_{Y_k} : \ \|s_{Y_k}\|_2 \le \Delta_k\}$$

and set $s_k = D_k Y_k s_{Y_k}$. The solution to (1.14) is of negligible cost once the matrices are formed, provided $|\mathcal{S}_k|$ is small.

[6] Matrix J^v is a diagonal matrix with each diagonal component equal to zero or unity. For example, if all the components of u and v are finite then $J^v = I$. If variable x_i has a finite lower bound and an infinite upper bound (or vice-versa) then strictly speaking v_i is not differentiable at a point $g_i = 0$; we define $J^v_{ii} = 0$ at such a point. Note that v_i is discontinuous at such a point but $v_i \cdot g_i$ is continuous.

[7] Notation: If z is a vector then $|z|^{\frac{1}{2}}$ denotes a vector with the i^{th} component equal to $|z_i|^{\frac{1}{2}}$.

[8] If A is a matrix then $< A >$ denotes the space spanned by the columns of A.

Note that if $\bar{M}_k$ is positive definite and the constraint $\|s_{Y_k}\|_2 \le \Delta_k$ is inactive then the solution to the reduced trust region problem is $s_k^N = D_k \bar{s}_k^N$ where

$$\bar{s}_k^N = -\bar{M}_k^{-1} \bar{g}_k. \tag{1.15}$$

In a neighbourhood of nondegenerate point satisfying second-order sufficiency, s_k^N is a Newton step for system (1.7).

Finally, we remark that many of the basic ideas behind the reflective Newton approach originated in previous work on various convex optimization problems, [6, 7, 8, 9, 21]. Note that convexity is not required in the new reflective Newton approach.

2. The Line Search. It is well known that a descent direction algorithm demands sufficient decrease at every step in order to achieve reasonable convergence properties. In the unconstrained setting, min $f(x)$, several such sufficiency conditions have been proposed. For example, Goldfarb [17] uses the modified Armijo[1] and Goldstein[18] conditions: Given $0 < \sigma_l < \sigma_u < 1$ and a descent direction s_k with $x_{k+1} = x_k + \alpha_k s_k$, α_k satisfies the modified Armijo/Goldstein conditions if

$$f(x_{k+1}) < f(x_k) + \sigma_l(\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 \min(s_k^T H_k s_k, 0)) \tag{2.1}$$

and

$$f(x_{k+1}) > f(x_k) + \sigma_u(\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 \min(s_k^T H_k s_k, 0)). \tag{2.2}$$

Roughly speaking condition (2.1) can be interpreted as restricting the step length from being too large relative to the decrease in f; condition (2.2) can be interpreted as restricting the step length from being relatively too small. Both conditions can be combined to form a single expression: If we define

$$\psi_k(\alpha) = \frac{f(x_{k+1}) - f(x_k)}{\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 \min(s_k^T H_k s_k, 0)} \tag{2.3}$$

conditions (2.1) and (2.2) can be expressed as

$$\sigma_l < \psi_k(\alpha_k) < \sigma_u. \tag{2.4}$$

We use conditions (2.1) and (2.2) for the piecewise linear path minimization process where $x_{k+1} = x_k + p_k(\alpha_k)$ and p_k is defined by (1.4).

Next we establish that there is an interval (α_l, α_u), depending on k, such that for all $\alpha \in (\alpha_l, \alpha_u)$, (2.4) is satisfied.

THEOREM 1. *Assume that $f(x)$ has two continuous derivatives and either $g_k^T s_k < 0$ or $g_k^T s_k = 0$ and $s_k^T H_k s_k < 0$ where $x_k \in int(\mathcal{F})$. Then either f is unbounded below along the piecewise linear path $p_k(\alpha)$ or, for $0 < \sigma_l < \sigma_u < 1$, there exists an interval (α_l, α_u), depending on k, such that condition (2.4) is satisfied.*

Proof. First we note that $\lim_{\alpha \to 0} \psi_k(\alpha) = 1$. To see this consider that from Taylor's theorem, for $\alpha < \beta_k^1$,

$$\psi_k(\alpha) = \frac{\alpha g_k^T s_k + \frac{1}{2}\alpha^2 s_k^T \bar{H}_k s_k}{\alpha g_k^T s_k + \frac{1}{2}\alpha^2 \min(s_k^T H_k s_k, 0)},$$

where

$$\bar{H}_k = H(x_k + \theta(\alpha)\alpha s_k), \quad 0 \le \theta(\alpha) \le 1.$$

Therefore, if $g_k^T s_k \neq 0$, $\psi_k(0) \stackrel{def}{=} \lim_{\alpha \to 0} \psi_k(\alpha) = 1$ and so $\psi_k(0) > \sigma_u > \sigma_l$; if $g_k^T s_k = 0$ then $s_k^T H_k s_k < 0$ and clearly $\psi_k(0) \stackrel{def}{=} \lim_{\alpha \to 0} \psi_k(\alpha) = 1$ and so $\psi_k(0) \stackrel{def}{=} \lim_{\alpha \to 0} \psi_k(\alpha) = 1$.

Assume $\psi_k(\alpha) \le \sigma_l$ for some $\alpha > 0$. Let α_u be the smallest α such that $\psi_k(\alpha) = \sigma_l$. Since $\psi_k(0) > \sigma_u > \sigma_l$ it follows that $\psi_k(\alpha) > \sigma_l$ for all $\alpha \in (0, \alpha_u)$. Therefore by continuity there exists a positive $\alpha_l < \alpha_u$ such that $\psi_k(\alpha) < \sigma_u$ for all $\alpha \in (\alpha_l, \alpha_u)$. Therefore (2.4) is satisfied on (α_l, α_u).

Now assume the contrary; i.e., $\psi_k(\alpha) > \alpha_l$ for all positive α. But since either $g_k^T s_k < 0$ or $s_k^T g_k = 0$ and $s_k^T H_k s_k < 0$, it follows that

$$\lim_{\alpha \to \infty} \alpha g_k^T s_k + \frac{1}{2}\alpha^2 \min(s_k^T H_k s_k, 0) = -\infty.$$

Therefore to achieve $\psi_k(\alpha) > \alpha_l$, for all positive α, it must be that

$$\lim_{\alpha \to \infty} f(x_k + p_k(\alpha)) - f(x_k) = -\infty.$$

Consequently f is unbounded below along the path $p_k(\alpha)$ as $\alpha \to \infty$. ∎

The interval (α_l, α_u) contains a finite number of breakpoints. Consequently, we can choose $\alpha_k \in (\alpha_l, \alpha_u)$ such that α_k is not a breakpoint.

A basic reflective path algorithm can now be stated. To allow for flexibility, especially with regard to the Newton step, we do not always require that both (2.1) and (2.2) be satisfied. Instead, we demand that either both these conditions are satisfied or (2.1) is satisfied and α_k is guaranteed to be bounded away from zero, e.g., $\alpha_k > \rho > 0$. The latter conditions are used to allow for the liberal use of Newton steps and do not weaken the global convergence results.

Note that since $x_1 \in int(\mathcal{F})$, it follows that $x_k \in int(\mathcal{F})$.

Algorithm 4 [ρ is a positive scalar.]

Choose $x_1 \in int(\mathcal{F})$.

For $k = 1, 2, ...$

1. Determine an initial descent dir'n s_k for f at x_k . Note that the piecewise linear path p_k is defined by x_k, s_k.
2. Perform an approximate piecewise line minimization of $f(x_k + p_k(\alpha))$, with respect to α, to determine α_k such that:
   (a) α_k does not correspond to a breakpoint
   (b) condition (2.1) is satisfied
   (c) Either
      i. α_k satisfies condition (2.2), *or*
      ii. $\alpha_k > \rho > 0$
3. $x_{k+1} = x_k + p_k(\alpha_k)$.

FIG. 8. *A reflective path algorithm satisfying line search conditions*

3. Constraint Compatibility and Consistency. Satisfaction of the piecewise line search condition in Algorithm 4 is not sufficient to ensure convergence. However, it turns out that this condition along with two restrictions on the descent direction s_k, "constraint-compatibility" and "consistency", are enough to obtain first-order convergence, i.e., to guarantee that $\{D_k^2 g_k\} \to 0$.

We begin with a discussion of constraint-compatibility.

DEFINITION 2. *A sequence of vectors* $\{w_k\}$ *is* **constraint-compatible** *if the sequence* $\{D_k^{-2} w_k\}$ *is bounded.* [9]

Constraint-compatibility of $\{s_k\}$ is important because it facilitates a sufficiently long step along s_k. In particular, if x_k is close to a boundary then a direction satisfying only $g_k^T s_k < 0$ may not guarantee that a sufficiently long step can be taken to obtain a convergence result – s_k may point directly at a nearby constraint and descent beyond this first breakpoint, along p_k, is not guaranteed. (Conditions (2.1) and (2.2) can still be satisfied though.) Constraint-compatibility helps avoid this problem by ensuring that the distances to breakpoints (corresponding to "correct sign conditions") remain bounded away from zero. Specifically, if $\{s_k\}$ is constraint-compatible then the positive distance to constraint j along s_k, $BR_k(j) = \max\{\frac{l_j - x_{k_j}}{s_{k_j}}, \frac{u_j - x_{k_j}}{s_{k_j}}\}$, is bounded away from zero for any j with the correct "sign condition". The "sign condition" refers to a consistency between v_j and $\max\{\frac{l_j - x_{k_j}}{s_{k_j}}, \frac{u_j - x_{k_j}}{s_{k_j}}\}$. The "sign condition" holds when $s_{k_j} g_{k_j} < 0$, and so $BR_k(j) = \frac{|v_{k_j}|}{|s_{k_j}|}$.

[9] Recall that the diagonal matrix D_k is defined by (1.11), i.e., $D_k^2 = D^2(x_k) = \mathrm{diag}(|v_k|)$

THEOREM 2. *If* $\{s_k\}$ *is a constraint-compatible sequence then* $\{BR_k(j) : BR_k(j) = \frac{|v_{k_j}|}{|s_{k_j}|}\}$ *is bounded away from zero.*

Proof. By constraint compatibility there exists $\rho > 0$ such that, for all iterations k and all indices j,

$$\frac{|s_{k_j}|}{|v_{k_j}|} \leq \rho.$$

Clearly if $BR_k(j) = \frac{|v_{k_j}|}{|s_{k_j}|}$, then $BR_k(j) \geq \frac{1}{\rho}$. ∎

Theorem 4 below establishes that several useful directions satisfy the constraint compatibility requirement. A technical lemma, and a compactness assumption, are required before stating and proving Theorem 4.

LEMMA 3. *Let* $\{s_k\}$ *be a sequence of vectors and assume* $\{s_k\}$ *is bounded. Assume that for each iteration* k *and each index* i *such that* $0 < |v_{k_i}| < 1$,

$$e_{k_i} s_{k_i} = |v_{k_i}| z_{k_i}, \tag{3.1}$$

where e_{k_i} *satisfies* $|e_{k_i}| \geq g_{k_i}^+$. *Assume* $\{z_k\}$ *is bounded. Then* $\{s_k\}$ *is constraint-compatible.*

Proof. Consider any subsequence, denoted by indices $\bar{k}$. If $\{v_{\bar{k}_i}\}$ is bounded away from zero then $\{\frac{s_{\bar{k}_i}}{|v_{\bar{k}_i}|}\}$ is bounded since, by assumption, $\{s_k\}$ is bounded. On the other hand, if $\{v_{\bar{k}_i}\} \to 0$ then by (1.10), $|e_{\bar{k}_i}| \geq \tau_g > 0$. But $\{z_{k_i}\} = \{\frac{e_{k_i} s_{k_i}}{|v_{k_i}|}\}$ is bounded by assumption; therefore, $\{\frac{s_{\bar{k}_i}}{|v_{\bar{k}_i}|}\}$ is bounded. Since every subsequence of $\{\frac{s_{k_i}}{|v_{k_i}|}\}$ is bounded, the sequence itself is bounded. ∎

Compactness and Smoothness Assumption: Given initial point $x_1 \in \mathcal{F}$, it is assumed that the level set $\mathcal{L} = \{x : x \in \mathcal{F} \text{ and } f(x) \leq f(x_1)\}$ is compact. Moreover, we assume $f(x)$ is twice continuously-differentiable on an open set $D \supseteq \mathcal{F}$.

THEOREM 4. *Assume* $0 < \Delta_l \leq \Delta_k \leq \Delta_u < \infty$, *where* Δ_l *and* Δ_u *are positive scalars satisfying* $\Delta_l < \Delta_u$. *Under the compactness and smoothness assumption, the following definitions yield constraint-compatible sequences* $\{s_k\}$*:*

1. $s_k = -D_k^2 g_k$
2. $s_k = -D_k^2 sgn(g_k)$ [10]

[10] If z is a vector then $w = \text{sgn}(z)$ is a vector: $w_i = 1$ if $z_i \geq 0$, $w_i = -1$ if $z_i < 0$.

3. $s_k = D_k u_k$, *where* u_k *is a unit eigenvector of* $\bar{M}_k$ *corresponding to a non-positive eigenvalue*
4. $s_k = D_k \bar{s}_k^N$ *where* $\bar{s}_k^N$ *is the Newton step in the scaled space,* $\bar{s}_k^N = -\bar{M}_k^{-1}\bar{g}_k$, *where* $\bar{g}_k = D_k g_k$ *and assuming* $\|\bar{s}_k^N\| \le \Delta_k \le \Delta_u$ *and* $\bar{M}_k$ *positive definite*
5. $s_k = \frac{D_k \bar{s}_k^N}{\|\bar{s}_k^N\|}$ *and assuming* $\|\bar{s}_k^N\| \ge \Delta_k \ge \Delta_l$ *and* $\bar{M}_k$ *positive definite*
6. s_k *is the solution to (1.8) with* $\mathcal{S}_k = \mathcal{R}^n$.

Proof. Constraint-compatibility of the first two choices for s_k follows directly from the definition and boundedness of $\{g_k\}$.

For case *3*, let (μ_k, u_k) be an eigenpair of $\bar{M}_k$ with $\mu_k \le 0$. Then

$$(\mu_k I - J_k^v D_k^{g^+}) s_k = D_k^2 H_k D_k u_k, \quad \mu_k \le 0,$$

where $D_k^{g^+} = \mathrm{diag}(g_k^+)$. For each index i let $\bar{k}_i$ denote the indices of any subsequence such that $|v_{\bar{k}_i}| < 1$. Then $J_{\bar{k}_i}^v = 1$ and $|\mu_k \bar{k}_i I - J_{\bar{k}_i}^v D_k^{g^+}| \ge g_{\bar{k}_i}^+$. Using compactness, $\{H_{\bar{k}} D_{\bar{k}} u_{\bar{k}}\}$ and $\{s_{\bar{k}}\} = \{D_{\bar{k}} u_{\bar{k}}\}$ are bounded. Therefore, by Lemma 3, $\{s_k\}$ is constraint-compatible.

For case *4*, note that s_k satisfies

$$J_k^v D_k^{g^+} s_k = -D_k^2 (g_k + H_k D_k \bar{s}_k^N).$$

But if $\|\bar{s}_k^N\| \le \Delta_k \le \Delta_u$ then, using compactness, both $\{g_k + H_k D_k \bar{s}_k^N\}$ and $\{s_k\}$ are bounded. Constraint-compatibility then follows from Lemma 3.

In case *5*,

$$J_k^v D_k^{g^+} s_k = -D_k^2 \left(\frac{g_k}{\|\bar{s}_k^N\|} + \frac{H_k D_k \bar{s}_k^N}{\|\bar{s}_k^N\|}\right).$$

But $\|\bar{s}_k^N\| \ge \Delta_k \ge \Delta_l > 0$; therefore, using compactness, $\{\frac{g_k}{\|\bar{s}_k^N\|} + \frac{H_k D_k \bar{s}_k^N}{\|\bar{s}_k^N\|}\}$ is bounded. The sequence $\{s_k\}$ is bounded since $s_k = \frac{D_k \bar{s}_k^N}{\|\bar{s}_k^N\|}$; constraint-compatibility follows from Lemma 3.

Finally in case *6* note that s_k satisfies

$$(J_k^v D_k^{g^+} + \mu_k I) s_k = -D_k^2 (g_k + H_k D_k \bar{s}_k) \tag{3.2}$$

for some $\mu_k \ge 0$ and $\bar{s}_k = D_k^{-1} s_k$. But $\|\bar{s}_k\| \le \Delta_k \le \Delta_u$ and so, using compactness, both $\{g_k + H_k D_k \bar{s}_k\}$ and $\{s_k\}$ are bounded. Therefore, Lemma 3 can be applied to yield constraint-compatibility. ∎

Note that a constraint-compatible sequence $\{s_k\}$ can be obtained by mixing the various steps s_k given in Theorem 4.

Constraint-compatibility is not sufficient to guarantee convergence. It is also important that first-order descent, represented by $g_k^T s_k$, be *consistent* with first-order optimality, represented by $D_k^2 g_k$. The following definition captures this concept.

DEFINITION 3. *A sequence* $\{w_k\}$ *satisfies the* **consistency condition** *if* $\{w_k^T g_k\} \to 0$ *implies* $\{D_k g_k\} \to 0$.

In Theorem 5 we give five useful examples of sequences that satisfy consistency.

THEOREM 5. *Under the compactness and smoothness assumption, the following definitions yield sequences* $\{s_k\}$ *satisfying the consistency condition.*

1. $s_k = -D_k^2 g_k$
2. $s_k = -D_k^2 sgn(g_k)$
3. $s_k = D_k \bar{s}_k^N$ *where* $\bar{s}_k^N = -\bar{M}_k^{-1}\bar{g}_k$, *assuming* $\bar{M}_k$ *is symmetric positive definite*
4. s_k *is a solution to (1.8) with* $\mathcal{S}_k = \mathcal{R}^n$.
5. s_k *is a solution to (1.8) where* $\mathcal{S}_k$ *has the property that* $w_k = D_k \bar{w}_k \in \mathcal{S}_k$ *for some vector* $\bar{w}_k$ *such that* $\{\|\bar{w}_k\|\}$ *is bounded away from zero and* $\{w_k\}$ *is consistent, i.e.,* $\{w_k^T g_k\} \to 0$ *implies* $\{D_k g_k\} \to 0$.

Proof.

1. The first case is clear since $-s_k^T g_k = \|D_k g_k\|_2^2$.
2. In this case $s_k^T g_k = \text{sgn}(g_k)^T D_k^2 g_k = \|D_k |g_k|^{\frac{1}{2}}\|$, and so the result follows.
3. If s_k is the Newton step then
$$-g_k^T s_k = (D_k g_k)^T \bar{M}_k^{-1}(D_k g_k).$$
But by compactness $\bar{M}_k$ is bounded, i.e., there exists a finite bound ρ_M such that $\|\bar{M}_k\|_2 \le \rho_M$. Therefore, $-g_k^T s_k \ge \frac{1}{\rho_M}\|D_k g_k\|^2$. The result follows.
4. The solution to (1.8) satisfies $s_k = D_k \bar{s}_k$ where[11]
$$\bar{s}_k = -(\bar{M}_k + \mu_k I)^+ \bar{g}_k + \omega_k u_k^1$$
where u_k^1 is a unit eigenvector corresponding to the most negative eigenvalue of $\bar{M}_k$ and $\bar{g}_k^T u_k^1 = 0$. Using a trust region solution characterization, e.g., [28], the matrix $\bar{M}_k + \mu_k I$ is positive semi-definite and $\bar{g}_k \in range(\bar{M}_k + \mu_k I)$. Since $\Delta_k \ge \Delta_l > 0$, it follows that $\{\mu_k\}$ is bounded above. Therefore, using compactness, $\{\bar{M}_k + \mu_k I\}$ is bounded and so there exists a positive scalar τ_M such that
$$\|\bar{M}_k + \mu_k I\|_2 \le \tau_M.$$
Therefore,
$$-g_k^T s_k = (D_k g_k)^T (\bar{M}_k + \mu_k I)^+ (D_k g_k) \ge \frac{1}{\tau_M}\|D_k g_k\|^2$$
and the result follows.

[11] If A is a matrix then A^+ denotes the pseudo-inverse of A.

5. Let $\mathcal{S}_k =< V_k >$ for some full-column rank matrix V_k; let Y_k be an orthonormalization of the columns of $D_k^{-1}V_k$. Since $w_k \in \mathcal{S}_k$ we can assume, without loss of generality, that one of the columns of Y_k is $\frac{\bar{w}_k}{\|\bar{w}_k\|}$. We can write the solution to (1.8) as $s_k = D_k Y_k s_{Y_k}$, where

$$s_{Y_k} = -(Y_k^T \bar{M}_k Y_k + \mu_k I)^+ Y_k^T \bar{g}_k + \omega_k u_k^1$$

where u_k^1 is a unit eigenvector corresponding to the most negative eigenvalue of $Y_k^T \bar{M}_k Y_k$ and $(Y_k^T \bar{g}_k)^T u_k^1 = 0$. Using a trust region solution characterizion, e.g., [28], the matrix $Y_k^T \bar{M}_k Y_k + \mu_k I$ is positive semi-definite and $(Y_k^T \bar{g}_k) \in range(Y_k^T \bar{M}_k Y_k + \mu_k I)$. Since $\Delta_k \geq \Delta_l > 0$, it follows that $\{\mu_k\}$ is bounded above. Therefore, using compactness, $\{Y_k^T \bar{M}_k Y_k + \mu_k I\}$ is bounded and so there exists a positive scalar τ_M such that

$$\|Y_k^T \bar{M}_k Y_k + \mu_k I\|_2 \leq \tau_M.$$

Therefore,

$$-g_k^T s_k = (Y_k D_k g_k)^T (Y_k^T \bar{M}_k Y_k + \mu_k I)^+ Y_k^T (D_k g_k) \geq \frac{1}{\tau_M} \|Y_k^T D_k g_k\|^2$$

Therefore $\{s_k^T g_k\} \to 0$ implies $\{\|Y_k^T D_k g_k\|\} \to 0$. However, $\frac{\bar{w}_k}{\|\bar{w}_k\|}$ is a column of Y_k and $\{\|\bar{w}_k\|\}$ is bounded from zero. Therefore, $\{\|Y_k^T D_k g_k\|\} \to 0$ implies $\{w_k^T g_k\} \to 0$ which implies $\{D_k g_k\} \to 0$ since $\{w_k\}$ is consistent (by assumption).

■

4. First-order convergence of the reflective path algorithm. In this section we establish that constraint-compatibility and consistency allow the reflective path algorithm, Algorithm 4, to achieve first-order convergence. Recall that a feasible point x is a first-order point if and only if $D^2(x)g(x) = 0$ where D is defined by (1.11).

All results are under the Compactness and Smoothness Assumption (Section 3).

Before stating the main result of this section a technical result is needed which says that the change in f along p_k is primarily represented by the linear term $g_k^T s_k$ as $\alpha_k \to 0$.

LEMMA 6. *Assume that $\{x_k\}$ is generated by the reflective path algorithm, Algorithm 4. Let $\{s_k\}$ be a sequence satisfying the consistency and constraint-compatibility conditions. Assume $\{\alpha_k\} \to 0$. Then,*

$$f(x_{k+1}) - f(x_k) = \alpha_k g_k^T s_k + O(\alpha_k^2).$$

Proof. Observe that if $0 < \beta_k^i < \alpha_k$ corresponding to variable x_j, then $s_{k_j} g_{k_j} \geq 0$ (from Theorem 2 and $\{\alpha_k\} \to 0$), where β_k^i is defined by Algorithm 2.

Without loss of generality, and for notational simplicity, suppose that the ordering of the breakpoints along s_k corresponds to the natural variable ordering. Note that since $\{\alpha_k\} \to 0$ we can assume that the indices corresponding to $0 < \beta_k^i < \alpha_k$ are distinct and so $\beta_k^i = BR_k(i)$ where BR is defined by (1.3). Assume that

$$0 \leq \beta_k^j < \alpha_k < \beta_k^{t_k+1}, \quad j = 1 : t_k. \tag{4.1}$$

Therefore,

$$s_{k_j} g_{k_j} \geq 0, \quad j = 1 : t_k. \tag{4.2}$$

By definition of the piecewise linear path p_k (see Algorithm 4) and using (4.2),

$$g_k^T s_k \geq g_k^T p_k^j, \quad j = 1 : t_k + 1. \tag{4.3}$$

Now using the definition of the breakpoints b_j^k (Algorithm 2) and applying Taylor's theorem (repeatedly),

$$\begin{aligned}
& f(x_{k+1}) - f(x_k) \\
= \;& f(x_{k+1}) - f(b_k^{t_k}) + \sum_{j=2}^{t_k} [f(b_k^j) - f(b_k^{j-1})] \;+\; f(b_k^1) - f(x_k) \\
= \;& (\alpha_k - \beta_k^{t_k}) \nabla f(b_k^{t_k})^T p_k^{t_k+1} + \sum_{j=2}^{t_k} [\beta_k^j - \beta_k^{j-1}] \nabla f(b_k^{j-1})^T p_k^j + \beta_k^1 \nabla f(x_k)^T p_k^1 \;+\; O(\alpha_k^2) \\
= \;& (\alpha_k - \beta_k^{t_k}) g_k^T p_k^{t_k+1} \;+\; \sum_{j=2}^{t_k} [\beta_k^j - \beta_k^{j-1}] g_k^T p_k^j \;+\; \beta_k^1 g_k^T p_k^1 \;+\; O(\alpha_k^2).
\end{aligned}$$

Now apply (4.3) to get

$$\begin{aligned}
f(x_{k+1}) - f(x_k) \;&\leq\; (\alpha_k - \beta_k^{t_k}) g_k^T s_k \;+\; \sum_{j=2}^{t_k} [\beta_k^j - \beta_k^{j-1}] g_k^T s_k \;+\; \beta_k^1 g_k^T s_k \;+\; O(\alpha_k^2) \\
&=\; \alpha_k g_k^T s_k \;+\; O(\alpha_k^2).
\end{aligned}$$

■

The main result in this section, first-order convergence, i.e., $\{D_k^2 g_k\} \to 0$, follows. Theorem 7 also establishes that $\{\alpha_k^2 \min(s_k^T H_k s_k, 0)\} \to 0$; this is not part of the first-order conditions but is useful subsequently.

THEOREM 7. *Assume that $\{x_k\}$ is a sequence generated by the reflective path algorithm (Algorithm 4) and that $\{s_k\}$ is the corresponding sequence satisfying both the consistency and constraint-compatibility conditions. Then the corresponding sequences $\{D_k^2 g_k\}$ and $\{\alpha_k^2 \min(s_k^T H_k s_k, 0)\}$ converge to zero.*

Proof. Since condition (2.1) is satisfied,

$$
\begin{aligned}
f(x_m) - f(x_0) &= \sum_{k=0}^{m-1} (f(x_{k+1}) - f(x_k)) \\
&< \sum_{k=0}^{m-1} (\sigma_l \alpha_k g_k^T s_k + \frac{1}{2}\sigma_l \alpha_k^2 \min(s_k^T H_k s_k, 0)) \\
&\leq 0.
\end{aligned}
$$

By the compactness and smoothness assumption, $\{f(x)\}$ is bounded on $\mathcal{F}$; therefore,

$$\lim_{k\to\infty} (\sigma_l \alpha_k g_k^T s_k + \frac{1}{2}\sigma_l \alpha_k^2 \min(s_k^T H_k s_k, 0)) = 0.$$

But

$$\sigma_l \alpha_k g_k^T s_k \leq 0 \quad \text{and} \quad \sigma_l \alpha_k^2 \min(s_k^T H_k s_k, 0) \leq 0$$

and so

$$\lim_{k\to\infty} \alpha_k g_k^T s_k = 0 \quad \text{and} \quad \lim_{k\to\infty} \alpha_k^2 \min(s_k^T H_k s_k, 0) = 0.$$

Now we establish that $\{D_k^2 g_k\}$ converges to zero by contradiction. Suppose this is not true. Since $\{s_k\}$ satisfies the consistency condition, $\{g_k^T s_k\}$ does not converge to zero. Hence $g_k^T s_k < -c$ for some $c > 0$. Therefore, $\{\alpha_k\}$ converges to zero. Using Lemma 6,

$$
\begin{aligned}
\lim_{k\to\infty} \psi_k(\alpha_k) &= \lim_{k\to\infty} \frac{f(x_{k+1}) - f(x_k)}{\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 \min(s_k^T H_k s_k, 0)} \\
&\geq \lim_{k\to\infty} \frac{\alpha_k g_k^T s_k + O(\alpha_k^2)}{\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 \min(s_k^T H_k s_k, 0)} \\
&= 1.
\end{aligned}
$$

This contradicts (2.2); hence, $\{D_k^2 g_k\}$ converges to zero. ∎

Theorems 4 and 5 provide several examples of directions satisfying consistency and constraint-compatibility; therefore, by Theorem 7, Algorithm 4 achieves first-order convergence with these choices.

5. Second-order convergence. In order to achieve a second-order algorithm (i.e., guarantee convergence to a second-order point; obtain quadratic convergence) we further specify the reflective path algorithm (Algorithm 4). In particular, we now assume that when $\bar{M}_k$ is positive definite and $\|\bar{s}_k^N\| \leq \Delta_k$ then the Newton step $s_k = D_k \bar{s}_k^N$ is taken; if $\bar{M}_k$ is not positive definite the direction s_k is defined by a reduced trust region problem [12]: s_k solves

$$\min_s \{s^T g_k + \frac{1}{2} s^T M_k s : \ \|D_k^{-1} s\|_2 \leq \Delta_k, \ s \in \mathcal{S}_k\}. \tag{5.1}$$

Algorithm 5

Choose $x_1 \in int(\mathcal{F})$.

For $k = 1, 2, ...,$

1. Determine initial descent dir'n s_k for f at x_k: If $\bar{M}_k$ is positive definite and $\|\bar{s}_k^N\| \leq \Delta_k$, choose $s_k = D_k \bar{s}_k^N$. If $\bar{M}_k$ is not positive definite choose $\Delta_k \in [\Delta_l, \Delta_u]$, choose subspace $\mathcal{S}_k$, and solve (5.1) to get s_k.
2. Determine α_k: *If* $s_k = s_k^N$ and $x_k + p_k(1)$ satisfies (2.1), then set $\alpha_k = 1$; *otherwise*, perform an approximate piecewise line minimization of $f(x_k + p_k(\alpha))$, with respect to α, to determine α_k such that
   (a) α_k is not a breakpoint;
   (b) α_k satisfies (2.1) and (2.2).
3. $x_{k+1} = x_k + p_k(\alpha_k)$.

FIG. 9. *A second-order reflective path algorithm*

Algorithm 5 presents a second-order reflective path algorithm.

Note: If $\alpha_k = 1$ is accepted by the line search but corresponds to a breakpoint, then modify α_k: $\alpha_k = \tilde{\alpha}_k \stackrel{def}{=} 1 - \epsilon_k$ where $\tilde{\alpha}_k$ is not a breakpoint, $\tilde{\alpha}_k$ satisfies (2.1), and $\epsilon_k < \chi_\alpha \|D_k g_k\|$ for some $\chi_\alpha > 0$.

The first important result of this section, Theorem 9, is that provided $\mathcal{S}_k$ is chosen so that negative curvature of $\bar{M}_k$ is "well-represented", Algorithm 5 generates points $\{x_k\}$ such that the second-order necessary conditions are satisfied at every limit point of $\{x_k\}$.

All results in the remainder of this paper are under the Compactness and Smoothness Assumption (Section 3).

A preliminary technical result is required. We denote the smallest eigenvalue of a real symmetric matrix A by $\lambda_{\min}(A)$. So if $\lambda(A) = \{\lambda_1, \lambda_2, ..., \lambda_n\}$, with $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$, then $\lambda_{\min}(A) = \lambda_1$.

LEMMA 8. *Assume that $\{x_k\}$ is generated by the second-order reflective path algorithm, Algorithm 5, where the initial point is strictly feasible. Let $\{s_k\}$ satisfy the consistency and constraint-compatibility conditions. Let $\mathcal{S}_k =< Y_k >$, for some orthonormal matrix Y_k, be chosen such that when $\lambda_{\min}(\bar{M}_k) \leq 0$,*

$$\lambda_{\min}(Y_k^T \bar{M}_k Y_k) \leq \max(-\epsilon_{nc}, \tau \lambda_{\min}(\bar{M}_k)), \tag{5.2}$$

for some $\epsilon_{nc} > 0$, $\tau > 0$. Then for any subsequence satisfying $\{\min(s_k^T H_k s_k, 0)\} \to 0$, the corresponding subsequence satisfies $\lim_{k\to\infty}\{\min(\lambda_{\min}(\bar{M}_k), 0)\} = 0$.

Proof. In this proof subscript k is identified with the subsequence under consider-

[12] We do not (yet) specify how s_k might be determined when $\bar{M}_k$ is positive definite and $\|\bar{s}_k^N\| > \Delta_k$.

ation. By definition, s_k satisfies

$$s_k^T H_k s_k + s_{Y_k}^T Y_k^T D_k^{g^+} Y_k s_{Y_k} + \mu_k \|s_{Y_k}\|^2 = s_{Y_k}^T Y_k^T D_k g_k.$$

But by Theorem 7 $\lim_{k\to\infty} D_k g_k = 0$, by assumption $\lim_{k\to\infty}\{\min(s_k^T H_k s_k, 0)\} = 0$, and $D_k^{g^+}$ is positive semidefinite. Moreover, since s_k solves (5.1), $\|s_{Y_k}\| = \Delta_k \geq \Delta_l > 0$; therefore,

$$\lim_{k\to\infty}\{\mu_k\} = 0.$$

However,

$$0 \leq -\min(\lambda_{\min}(Y_k^T \bar{M}_k Y_k), 0) \leq \mu_k,$$

hence

$$\lim_{k\to\infty}\{\min(\lambda_{\min}(Y_k^T \bar{M}_k Y_k), 0)\} = 0,$$

and applying assumption (5.2),

$$\lim_{k\to\infty}\{\min(\max(-\epsilon_{nc}, \tau\lambda_{\min}(\bar{M}_k)), 0)\} = 0.$$

Hence

$$\lim_{k\to\infty}\{\min(\lambda_{\min}(\bar{M}_k), 0)\} = 0$$

■

THEOREM 9. *Assume that x_* is a nondegenerate limit point of $\{x_k\}$. If the assumptions of Lemma 8 hold then $\lambda_{\min}(\bar{M}_*) \geq 0$.*

Proof. Our proof is by contradiction. Assume

$$\lambda_{\min}(\bar{M}_*) < 0.$$

Applying Lemma 8, this means that there exists a subsequence with

$$\lim_{k\to\infty} \min(s_k^T H_k s_k, 0) < 0.$$

Using Theorem 7, $\lim_{k\to\infty} \alpha_k \min(s_k^T H_k s_k, 0) = 0$; hence, $\lim_{k\to\infty} \alpha_k = 0$.

By Theorem 7, $D_* g_* = 0$, and by assumption, x_* is a nondegenerate point; therefore, for k sufficiently large, $\mathrm{sgn}(g_{k_j}) = \mathrm{sgn}(g_{*_j})$ if $j \notin Free_*$. Hence, for any $j \notin Free_*$, $BR_k(j) = \frac{|v_{k_j}|}{|s_{k_j}|}$. Alternatively, if $j \in Free_*$ then $|BR_k(j)| \to \infty$. By Theorem 2, $\{BR_k(j) : BR_k(j) = \frac{|v_{k_j}|}{|s_{k_j}|}\}$ is bounded away from zero. It follows, since $\alpha_k \to 0$, that $0 \leq \alpha_k < \beta_k^1$ for sufficiently large k, where β is defined by Algorithm 2. Therefore, due to the absence of breakpoints on $(0, \alpha_k)$, Taylor's Theorem can be applied straightforwardly to yield, for some subsequence:

$$\begin{aligned}
\lim_{k\to\infty} \psi_k(\alpha_k) &= \lim_{k\to\infty} \frac{f(x_k + \alpha_k s_k) - f(x_k)}{\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 \min(s_k^T H_k s_k, 0)} \\
&= \lim_{k\to\infty} \frac{\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 s_k^T H(x_k + \theta(\alpha_k)) s_k}{\alpha_k g_k^T s_k + \frac{1}{2}\alpha_k^2 s_k^T H_k s_k}, \qquad 0 \le \theta(\alpha_k) \le \alpha_k \\
&= 1.
\end{aligned}$$

This contradicts condition (2.2). Hence we conclude that every nondegenerate limit point is a second order point. ■

Next we work toward establishing convergence of the entire sequence $\{x_k\}$.

First we establish that there is a natural (local) Newton process for problem (1.1). This view is similar to the development given in [6] for the convex quadratic problem. Let x_* be a specified nondegenerate point satisfying the second-order sufficiency conditions.

Consider a finite set $\mathcal{V}$ of functions defined with respect to x_*:

$$F_\nu(x) = D_\nu(x) g(x) \tag{5.3}$$

where $D_\nu(x) = \mathrm{diag}(\nu(x))$ and $\nu(x)$ is a vector defined

$$\nu_i = \begin{cases} +1 \text{ or } -1 \text{ or } x_i - u_i \text{ or } x_i - l_i & \text{if } g_i^* = 0 \\ x_i - u_i & \text{if } g_i^* < 0 \\ x_i - l_i & \text{if } g_i^* > 0. \end{cases} \tag{5.4}$$

Note: When $g_i^* = 0$ the choice $\nu_i = x_i - u_i$ is valid only when u_i is finite; the choice $\nu_i = x_i - l_i$ is valid only when l_i is finite.

Each function F_ν is twice continuously differentiable; furthermore, $F_\nu(x_*) = 0$ for every possible ν. Of course F_ν cannot be used computationally since x_* is not known a priori. However, since each step of our proposed algorithms is an approximate Newton step for exactly one set of equations based on the definition of $v(x)$, i.e., $\nu(x) = v(x)$, $\mathcal{V}$ and F_ν are useful in a theoretical sense to help establish asymptotic convergence results of our proposed algorithm.

The next result formalizes the simple observation that any member of $\mathcal{V}$ can be used interchangeably with any other, at any iteration, and there remains a neighbourhood around x_* retaining quadratic convergence properties of a Newton process.

THEOREM 10. *Let $\mathcal{V} = \{F_\nu : R^n \to R^n\}$ be a finite set of functions satisfying the following assumptions:*

- *Each F_ν is continuously differentiable in an open convex set $\mathcal{C}$.*
- *There is a x_* in $\mathcal{C}$ such that $F_\nu(x_*) = 0$ and $\nabla F_\nu(x_*)$ is nonsingular for all $F_\nu \in \mathcal{S}$.*

- *There is a constant κ_0 such that for all $F_\nu \in \mathcal{S}$,*

$$\|\nabla F_\nu(x) - \nabla F_\nu(x_*)\| \leq \ \kappa_0 \|x - x_*\|, \tag{5.5}$$

  for $x \in \mathcal{C}$.

Let $\{x_k\}$ and $\{s_k\}$ be sequences such that $x_{k+1} = x_k + s_k$ and suppose

$$\|s_k - s_k^{N_{\nu_k}}\| = O(\|x_k - x_*\|)^2,$$

where $s_k^{N_{\nu_k}}$ is the Newton step for one of the function $F_{\nu_k} \in \mathcal{V}$ at x_k, i.e.,

$$s_k^{N_{\nu_k}} = -(\nabla F_{\nu_k}(x_k))^{-1} F_{\nu_k}(x_k).$$

Then, for $\mathcal{C}$ sufficiently small, $\{x_k\}$ converges quadratically to x^*.

Proof. The argument is straightforward and uses a standard result in the last step, e.g., [26],:

$$\begin{aligned}
\|x_{k+1} - x_*\| &= \|x_k + s_k - x_*\| \\
&= \|x_k + s_k^{N_{\nu_k}} - x_* + s_k - s_k^{N_{\nu_k}}\| \\
&\leq \|x_k + s_k^{N_{\nu_k}} - x_*\| + \|s_k - s_k^{N_{\nu_k}}\| \\
&= O(\|x_k - x_*\|)^2.
\end{aligned}$$

■

Our next main result is that the local reflective Newton method is locally and quadratically convergent. The Local Reflective Newton Method, given in Algorithm 6, is merely Algorithm 3 with direction s_k specified as the Newton step and α_k chosen so that $|\alpha_k - 1| = O(\|x_k - x_*\|)$. We assume that $x_1 \in int(\mathcal{F})$.

Algorithm 6

Choose $x_1 \in int(\mathcal{F})$.
For $k = 1, 2, ...,$
1. Solve $\bar{M}_k \bar{s}_k^N = -\bar{g}_k = D_k g_k$, set $s_k = D_k \bar{s}_k^N$.
2. Determine α_k s.t. $|\alpha_k - 1| = O(\|x_k - x_*\|)$ and $x_k + p_k(\alpha_k) \in int(\mathcal{F})$.
3. $x_{k+1} = x_k + p_k(\alpha_k)$.

FIG. 10. *A local reflective Newton method*

Note that the k^{th} iteration is computable provided x_k is sufficiently close to x_* and $x_k \neq x_*$. To see this note that the Newton direction and the step size α_k are always computable in a neighbourhood of x_*. In particular, $\bar{M}_k$ is positive definite in a neighbourhood of x_*, assuming x_* is nondegenerate and satisfies second-order sufficiency, and $\bar{g}_k \neq 0$ unless $x_k = x_*$. Stepsize $\alpha_k = 1$ satisfies the stepsize condition (step 2. in Algorithm 5) unless $x_k + p_k(1)$ is on the boundary, i.e., $(x_k + p_k(1))_j$ is tight for some index j. In this case α_k can be chosen slightly smaller than unity, satisfying $|\alpha_k - 1| = O(\|x_k - x_*\|)$, and strict feasibility will be maintained. Computationally, the condition $|\alpha_k - 1| = O(\|x_k - x_*\|)$ can be assured by using the facts that $\|D_k g_k\| = O(\|x_k - x_*\|)$ and $\|D_k g_k\|$ is computable at x_k.

A key observation is that, provided x_* satisfies nondegeneracy and second-order sufficiency and x_1 is sufficiently close to x_*, the search direction s_k generated by Algorithm 6 is a Newton step for one of the functions in $\mathcal{V}$. Therefore, to establish quadratic convergence we focus on the relationship between $p_k(\alpha_k)$ and s_k. The following result provides the necessary connection.

LEMMA 11. *Let x_* be a nondegenerate point satisfying second-order sufficiency conditions. Assume that $\nu(x)$ is chosen such that $\nu(x) = v(x)$. Let $s^N(x)$ be the corresponding Newton direction, i.e.,*

$$s^N(x) \;=\; -(D^2H + J^vD^g)^{-1}D^2g \tag{5.6}$$

where $g = g(x) = \nabla f(x)$, $H = H(x) = \nabla^2 f(x)$, $D^g = D^g(x) = diag(|g|)$, $D^2 = D^2(x) = diag(|v(x)|)$, $J^v = J^v(x)$ is the diagonal Jacobian[13] *matrix of v. There exists an open neighborhood $\mathcal{C}$ containing x_* such that for all $x \in int(\mathcal{F})\bigcap\mathcal{C}$, $s^N(x)$ is well-defined and for each $j \notin Free_*$,*

$$|1 - \beta_j^N(x)| = O(\|x_* - x\|) \tag{5.7}$$

where $\beta_j^N = \frac{|v_j(x)|}{|s_j^N(x)|}$.

Proof. Since x_* satisfies nondegeneracy and second-order sufficiency, it follows that the matrix $D^{(2)}H + J^vD^g$ is nonsingular in a neighbourhood of x_* and so $s^N(x)$ is well-defined. From the definition of the Newton step (5.6) it follows that if $j \notin Free_*$,

$$s_j^N = -|v_j| \cdot \mathrm{sgn}(g_j) - \frac{|v_j|}{|g_j|}(Hs^N)_j$$

which implies

$$|v_j| - \frac{|v_j|}{|g_j|} \cdot |(Hs^N)_j| \le |s_j^N| \le |v_j| + \frac{|v_j|}{|g_j|}|(Hs^N)_j|. \tag{5.8}$$

The first inequality in (5.8) uses the fact that $g_j^* \neq 0$ (by nondegeneracy), and $Hs^N \to 0$ as $x \to x^*$. Therefore,

$$1 - \frac{|(Hs^N)_j|}{|g_j|} \le \frac{|s_j^N|}{|v_j|} \le 1 + \frac{|(Hs^N)_j|}{|g_j|}. \tag{5.9}$$

[13] Matrix J^v is a diagonal matrix with each diagonal component equal to zero or unity. For example, if all the components of u and v are finite then $J^v = I$. If variable x_i has a finite lower bound and an infinite upper bound (or vice-versa) then strictly speaking v_i is not differentiable at a point $g_i = 0$; we define $J_{ii}^v = 0$ at such a point. Note that v_i is discontinuous at such a point but $v_i \cdot g_i$ is continuous.

But, by nondegeneracy and continuity, $|g_j|$ is bounded away from zero in a neighbourhood of x_*; H is bounded; $\|s^N\| = O(\|x - x_*\|)$; therefore, from (5.9) it is easy to show that $|1 - \beta_j^N| = O(\|x - x_*\|)$. ■

THEOREM 12. *Let x_* be a nondegenerate point satisfying the second-order sufficiency conditions. Assume that $\{x_k\}$ is generated by Algorithm 6. Then, for $x_1 \in int(\mathcal{F})$ and sufficiently close to x_*, $\{x_k\} \in int(\mathcal{F})$ and $\{x_k\}$ converges quadratically to x_*.*

Proof. Let β_k^1 be the steplength to the first breakpoint along direction s_k. If $\alpha_k < \beta_k^1$ then $p_k(\alpha_k) = \alpha_k s_k$ where s_k is the Newton step. However, $|\alpha_k - 1| = O(\|x_k - x_*\|)$ and since s_k is the Newton step for some function in $\mathcal{F}$, $\|s_k\| = O(\|x_k - x_*\|)$; therefore, $\|p_k(\alpha_k) - s_k\| = O(\|x_k - x_*\|^2)$ and so Theorem 10 applies and the result follows.

Assume that $\beta_k^{t_k} < \alpha_k < \beta_k^{t_k+1}$. From the definition of the reflective process, we can write

$$p_k(\alpha_k) - s_k = \sum_{i=2}^{t_k} (\beta_k^i - \beta_k^{i-1}) p_k^i + (\alpha_k - \beta_k^{t_k}) p_k^{t_k+1} + \beta_k^1 s_k - s_k.$$

But applying Lemma 11,

$$\|p_k(\alpha_k) - s_k\| = O(\|s_k\| \cdot \|x_k - x_*\|)$$

But s_k is the Newton step for some function in $\mathcal{F}$; hence, $\|s_k\| = O(\|x_k - x_*\|)$. It follows that $\|p_k(\alpha_k) - s_k\| = O(\|x_k - x_*\|^2)$; applying Lemma 10 the result follows. ■

We have established global convergence results for Algorithm 4 (and therefore Algorithm 5) and we have established that the local reflective Newton method, Algorithm 6, yields quadratic convergence. We now show that Algorithm 5 reduces to Algorithm 6 in a neighbourhood of a nondegenerate second-order point: global and quadratic convergence properties follow. In particular, we show that in a neighbourhood of a nondegenerate point satisfying second-order sufficiency conditions, a Newton step will satisfy line search condition (2.1).

THEOREM 13. *Assume x_* is a nondegenerate point satisfying second-order sufficiency conditions and τ_g is sufficiently small* [14]. *Let $0 < \sigma_l < \frac{1}{2}$. Suppose $\{x_k\}$ is generated by Algorithm 6. Then for x_1 sufficiently close to x_* and k sufficiently large,*

$$f(x_k + p_k(\alpha_k)) < f(x_k) + \sigma_l (g_k^T s_k + \frac{1}{2}\min(s_k^T H_k s_k, 0)). \tag{5.10}$$

Proof. Suppose there are $t_k - 1$ breakpoints $b_1, b_2, \ldots, b_{t_k-1}$, to the left of α_k, corresponding to step lengths $\beta_k^1, \beta_k^2, \ldots, \beta_k^{t_k-1}$. For notational simplicity let us label

[14] τ_g is used in the definition of the extended gradient (1.10).

$x_k + p_k(\alpha_k)$ with $b_k^{t_k}$. Clearly,

$$f(x_k + p_k(\alpha_k)) - f(x_k) = f(b_k^1) - f(x_k) + \sum_{i=1}^{t_k-1} [f(b_k^{i+1}) - f(b_k^i)]. \tag{5.11}$$

Note that $p_k^{i+1} = D_k^{\sigma_{i+1}} s_k$ where $D_k^{\sigma_{i+1}}$ is a diagonal matrix with each diagonal entry equal to ± 1; therefore $\|p_k^{i+1}\| = O(\|s_k\|)$. Consequently, applying Lemma 11, for any $1 \le i \le t_k - 1$,

$$\begin{aligned}
& f(b_k^{i+1}) - f(b_k^i) \\
= \; & (\beta_k^{i+1} - \beta_k^i) g(b_k^i)^T p_k^{i+1} + \frac{1}{2}(\beta_k^{i+1} - \beta_k^i)^2 (p_k^{i+1})^T H_k^i p_k^{i+1} + o(\|(\beta_k^{i+1} - \beta_k^i) p_k^{i+1}\|^2) \\
= \; & (\beta_k^{i+1} - \beta_k^i) g(b_k^i)^T p_k^{i+1} + \frac{1}{2}(\beta_k^{i+1} - \beta_k^i)^2 (p_k^{i+1})^T H_k^i p_k^{i+1} + o(\|s_k\|^2) \\
= \; & (\beta_k^{i+1} - \beta_k^i) g_k^T p_k^{i+1} + \frac{1}{2}(\beta_k^{i+1} - \beta_k^i)^2 (p_k^{i+1})^T H_k^i p_k^{i+1} + o(\|s_k\|^2) \\
= \; & (\beta_k^{i+1} - \beta_k^i) g_k^T p_k^{i+1} + \frac{1}{2}(\beta_k^{i+1} - \beta_k^i)^2 (s_k)^T D_k^{\sigma_{i+1}} H_k^i D_k^{\sigma_{i+1}} s_k + o(\|s_k\|^2) \\
= \; & (\beta_k^{i+1} - \beta_k^i) g_k^T p_k^{i+1} + o(\|s_k\|^2).
\end{aligned}$$

Moreover, using Taylor's theorem and Lemma 11,

$$\begin{aligned}
f(b_k^1) - f(x_k) &= \beta_k^1 g_k^T s_k + \frac{1}{2}(\beta_k^1)^2 s_k^T H_k s_k + o(\|s_k\|^2) \\
&= g_k^T s_k + \frac{1}{2} s_k^T H_k s_k + o(|g_k^T s_k|) + o(\|s_k\|^2).
\end{aligned}$$

The most difficult term to deal with is $g_k^T p_k^{i+1}$; however, we can show that $|g_k^T p_k^{i+1}| = O(-g_k^T s_k)$ and this leads the way to the final result. To show this we use the fact that, due to second-order sufficiency, there exists $\mu > 0$ such that for all k sufficiently large,

$$s_k^T M_k s_k \;\; \ge \;\; \mu \|s_k\|^2, \tag{5.12}$$

and

$$\bar{s}_k^T \bar{M}_k \bar{s}_k \ge \mu \|\bar{s}_k\|^2.$$

But since s_k is the Newton direction,

$$g_k = -M_k s_k = -D_k^{-1} \bar{M}_k D_k^{-1} s_k = -D_k^{-1} \bar{M}_k \bar{s}_k;$$

therefore,

$$-g_k^T s_k = \bar{s}_k^T \bar{M}_k \bar{s}_k \ge \mu \|\bar{s}_k\|^2. \tag{5.13}$$

But $p_k^{i+1} = D_k^{\sigma_{i+1}} s_k$ where $D_k^{\sigma_{i+1}}$ is a diagonal matrix with each diagonal element equal to ± 1. Hence, using the boundedness of $\{\bar{M}_k\}$,

$$|-g_k^T p_k^{i+1}| = |s_k^T D_k^{\sigma_{i+1}} M_k s_k| = |\bar{s}_k^T D_k^{\sigma_{i+1}} \bar{M}_k \bar{s}_k| = O(\|\bar{s}_k\|^2). \tag{5.14}$$

Therefore, combining (5.13) and (5.14),

$$|-g_k^T p_k^{i+1}| = O(-g_k^T s_k). \tag{5.15}$$

Collecting together the terms above, and applying Lemma 11, (5.11) becomes

$$f(x_k + p_k(\alpha_k)) - f(x_k) = g_k^T s_k + \frac{1}{2} s_k^T H_k s_k + o(|g_k^T s_k|) + o(\|s_k\|^2).$$

But $-g_k^T s_k = s_k^T M_k s_k \geq \mu \|s_k\|^2$, from (5.12). Therefore,

$$\begin{aligned} f(x_k + p_k(\alpha_k)) - f(x_k) &= g_k^T s_k + \frac{1}{2} s_k^T H_k s_k + o(|g_k^T s_k|) \\ &= \frac{1}{2} g_k^T s_k - \frac{1}{2} s_k^T D_k^{\frac{g}{v}} s_k + o(|g_k^T s_k|). \end{aligned} \tag{5.16}$$

But, for k sufficiently large,

$$o(|g_k^T s_k|) \leq -\frac{(1-2\sigma_l)}{2} g_k^T s_k \tag{5.17}$$

and $-s_k^T D_k^{\frac{g}{v}} s_k \leq \min(s_k^T H_k s_k, 0)$ and so, using (5.16).

$$f(x_k + p_k(\alpha_k)) - f(x_k) < \sigma_l s_k^T g_k + \frac{1}{2} \min(s_k^T H_k s_k, 0)$$

which implies for $\sigma_l < 1$,

$$f(x_k + p_k(\alpha_k)) - f(x_k) < \sigma_l (s_k^T g_k + \frac{1}{2} \min(s_k^T H_k s_k, 0))$$

■

THEOREM 14. *Assume $\{x_k\}$ is generated by Algorithm 5 and τ_g is sufficiently small. Let $\{s_k\}$ satisfy constraint-compatibility and consistency. Suppose Y_k is a matrix with orthonormal columns and let $\mathcal{S}_k = < Y_k >$ be chosen such that, when $\lambda_{\min}(\bar{M}_k) \leq 0$,*

$$\lambda_{\min}(Y_k^T \bar{M}_k Y_k) \leq \max(-\epsilon_{nc}, \tau \lambda_{\min}(\bar{M}_k)), \tag{5.18}$$

for some $\epsilon_{nc} > 0$, $\tau > 0$. Then,

- *Every limit point of $\{x_k\}$ is a first-order point.*
- *Every nondegenerate limit point satisfies the second-order necessary conditions.*
- *If a nondegenerate limit point x_* satisfies second-order sufficiency conditions then, provided τ_g is sufficiently small, $\{x_k\}$ is convergent to x_*. The convergence rate is quadratic, i.e.,*

$$\|x_{k+1} - x_*\| = O(\|x_k - x_*\|^2).$$

Proof. By Theorems 7 and 9 every limit point satisfies the second-order necessary conditions. Let x_* be a limit point satisfying nondegeneracy and second-order sufficiency conditions. By Theorem 13 a unit step size[15], for some constant $\chi_\alpha > 0$. will satisfy (2.1) for $\|x_k - x_*\|$ sufficiently small. Therefore, for $\|x_k - x_*\|$ sufficiently small, Algorithm 5 reduces to Algorithm 6: quadratic convergence follows from Theorem 12. ■

Therefore if we determine s_k by solving (5.1) at each iteration with $\mathcal{S}_k = \mathcal{R}^n$, for example, then the assumptions of Theorem 14 will be satisfied and so second-order convergence will be attained. We state this formally.

COROLLARY 15. *Assume $x_1 \in int(\mathcal{F})$ and let $\{x_k\}$ be generated by Algorithm 5 with $\{s_k\}$ determined by solving (5.1) at each iteration with $\mathcal{S}_k = \mathcal{R}^n$. Then,*

- *Every limit point of $\{x_k\}$ is a first-order point.*
- *Every nondegenerate limit point satisfies the second-order necessary conditions.*
- *If a nondegenerate limit point x_* satisfies second-order sufficiency conditions then, provided τ_g is sufficiently small, $\{x_k\}$ is convergent to x_*; the convergence rate is quadratic, i.e.,*

$$\|x_{k+1} - x_*\| = O(\|x_k - x_*\|^2).$$

Proof. By Theorems 4 and 5 the sequence $\{s_k\}$ satisfies constraint-compatibility and consistency. Since (5.1) is used to define s_k with $\mathcal{S}_k = \Re^n$, it follows that condition (5.18) is satisfied. Therefore, the assumptions of Theorem 14 are satisfied and the result follows. ■

6. A Practical Reflective Newton Algorithm for Large-Scale Problems. Algorithm 5 allows for some freedom in the determination of the direction s_k. As we have already remarked, if we determine s_k by solving (5.1) at each iteration with $\mathcal{S}_k = R^n$, then second-order convergence ensues (Corollary 15). However, this choice can lead to expensive subproblems (5.1), especially when n is large. Therefore it is worthwhile exploring alternative choices for $\mathcal{S}_k$, particularly if we can maintain the strong convergence properties for small values of $|\mathcal{S}_k|$. Below we propose a specific way to choose $\mathcal{S}_k$, restricting $|\mathcal{S}_k| \leq 2$, whilst retaining strong second-order convergence properties.

Constraint-compatibility plays a key role in the convergence of a reflective path algorithm. If a reduced trust region problem (5.1) is used to solve for a direction s_k – which, in turn, defines the piecewise linear path p_k – the subspace $\mathcal{S}_k$ must be chosen with constraint-compatibility in mind. It is easy to see that if s_k solves (5.1) for some subspace $\mathcal{S}_k$ then $\{D_k^{-1}s_k\}$ is bounded. This observation leads to the following two technical results.

[15] If $\alpha_k = 1$ corresponds to a breakpoint then $\alpha_k = \tilde{\alpha}_k = 1 - \epsilon_k$ where $\tilde{\alpha}_k$ is not a breakpoint, $\tilde{\alpha}_k$ satisfies (2.1), and $\epsilon_k < \chi_\alpha\|D_k g_k\|$

LEMMA 16. *Let $\{Y_k\}$ be a sequence of matrices where each matrix Y_k has orthonormal columns and suppose $\mathcal{S}_k =< D_k Y_k >$. Assume every column of $D_k Y_k$ generates a constraint-compatible sequence. Let $u_k \in \mathcal{S}_k$; assume the sequence $\{D_k^{-1} u_k\}$ is bounded. Then, the sequence $\{u_k\}$ is constraint-compatible.*

Proof. If $u_k \in \mathcal{S}_k$ then $u_k = D_k Y_k w_k$ for some vector w_k. But $\{D_k^{-1} u_k\}$ is bounded by assumption; therefore, $\{Y_k w_k\}$ is bounded and, by orthonormality of the columns of Y_k, the sequence $\{w_k\}$ is bounded. It is now easy to see that $\{u_k\}$ is constraint-compatible, i.e., $\{D_k^{-2} u_k\}$ is bounded. To see this notice that the sequence generated by any column of $D_k^{-2}(D_k Y_k)$ is bounded, by assumption, and we have already argued that $\{w_k\}$ is bounded. Therefore, since $u_k = D_k Y_k w_k$, the result follows. ∎

In the next lemma we indicate that the application of Lemma 16 is straighforward in the 2-dimensional case – subsequently we will use it in this setting. A definition is needed.

Definition: Let $\mathcal{A}$ be a subspace and w a vector. Define $r(\mathcal{A}, w)$ to be the *residual vector* of the orthogonal projection of w onto $\mathcal{A}$. If the columns of matrix Y form an orthonormal basis for $\mathcal{A}$, then $r(\mathcal{A}, w) = w - YY^T w$.

LEMMA 17. *Let a_k be a unit vector and suppose the sequences $\{D_k a_k\}$ and $\{D_k b_k\}$ are constraint-compatible; assume there exists a constant $\tau > 0$ such that $r(a_k, b_k) > \tau$ for all k. Then if $u_k \in \mathcal{S}_k =< D_k a_k, D_k b_k >$ and $\{D_k^{-1} u_k\}$ is bounded, then $\{u_k\}$ is constraint-compatible.*

Proof. Let $y_k^1 = a_k$ and so

$$r(a_k, b_k) = b_k - [(y_k^1)^T b_k] y_k^1.$$

Since $\{D_k y_k^1\}$ and $\{D_k b_k\}$ are both constraint-compatible, and $\{b_k\}$ is bounded due to constraint-compatibility of $\{D_k b_k\}$, $\{D_k r_k\}$ is constraint-compatible. Let $y_k^2 = \frac{r_k}{\|r_k\|}$. From $\|r_k\| > \tau > 0$, $\{D_k y_k^2\}$ is constraint-compatible. Since $\{D_k y_k^1\}$ and $\{D_k y_k^2\}$ are constraint-compatible and $\{\mathcal{S}_k\} = \{< D_k a_k, D_k b_k >\} = \{< D_k Y_k >\}$, it follows from Lemma 16 that $\{s_k\}$ is constraint-compatible. ∎

The next algorithm, Algorithm 7, describes a particular way to choose s_k (and $\mathcal{S}_k$, when appropriate) with the large-scale setting in mind. Each subspace $\mathcal{S}_k$ satisfies $|\mathcal{S}_k| \leq 2$ and so problem (5.1) is inexpensive.

Two technical results pertaining to Algorithm 7 are needed before establishing the main theorem. Let ρ_M be the maximum spectral radius of $\bar{M}(x)$ on $\mathcal{L} = \{x : \ x \in$

Algorithm 7[Let $\tau < 1$, τ_1, and τ_2 be small positive constants.]

Case 0: $\bar{M}_k$ is positive definite and $\|\bar{s}_k^N\| \le \Delta_k$.
Set $s_k = s_k^N = -D_k \bar{M}_k^{-1} \bar{g}_k = D_k \bar{s}_k^N$.
Case 1: $\bar{M}_k$ is positive definite and $\|\bar{s}_k^N\| > \Delta_k$.

if $\|r(\bar{s}_k^N, \bar{g}_k)\| > \tau_1$
$\mathcal{S}_k =< D_k^2 g_k, s_k^N >$, solve (5.1) to get s_k.
else
set $s_k = -D_k^2 g_k$
end

Case 2: $\bar{M}_k$ is not positive definite. Compute $w_k = D_k \bar{w}_k$, where $\bar{w}_k$ is a unit vector such that $\{w_k\}$ is constraint-compatible and

$$\bar{w}_k^T \bar{M}_k \bar{w}_k \le \max\{-\epsilon_{nc}, \tau \lambda_{\min}(\bar{M}_k)\}$$

Let $\bar{z}_k = \frac{D_k \text{sgn}(g_k)}{\|D_k \text{sgn}(g_k)\|}$.

if $\|r(\bar{w}_k, \bar{z}_k)\| < \max(\|D_k g_k\|, -\tau_2 \bar{w}_k^T \bar{M}_k \bar{w}_k)$
$\mathcal{S}_k =< D_k^2 \text{sgn}(g_k) >$, solve (5.1) to get s_k.
else
$\mathcal{S}_k =< D_k^2 \text{sgn}(g_k), D_k \bar{w}_k >$, solve (5.1) to get s_k.
end

FIG. 11. *Determination of the descent direction s_k*

$\mathcal{F}$ and $f(x) \le f(x_1)\}$. Since $\rho(\bar{M}(x))$ is continuous on $\mathcal{L}$, a compact set, the upper bound ρ_M exists.

LEMMA 18. *Assume $\{x_k\}$ is generated by Algorithm 5 with $\{s_k\}$ generated by Algorithm 7. Then,*

1. *the subsequence $\{\|D_k sgn(g_k)\| : \lambda_{\min}(\bar{M}_k) < 0\}$ is bounded away from zero,*
2. *the subsequence $\{z_k = D_k \bar{z}_k : \lambda_{\min}(\bar{M}_k) < 0\}$ is constraint-compatible, where $\bar{z}_k = \frac{D_k sgn(g_k)}{\|D_k sgn(g_k)\|}$.*

Moreover, if we assume that $\tau_2 < \frac{1}{5\rho_M}$, and that corresponding to any subsequence $\{\mathcal{S}_k\} = \{< D_k^2 sgn(g_k) >\}$, $\{D_k g_k\}$ converges to zero, and $\lim_{k\to\infty} \lambda_{\min}(\bar{M}_k) < 0$, then

$$\bar{z}_k^T \bar{M}_k \bar{z}_k < \frac{1}{2} \bar{w}_k^T \bar{M}_k \bar{w}_k$$

for sufficiently large k.

Proof. First assume there exists a subsequence with $\lim_{k\to\infty}\{D_k \mathrm{sgn}(g_k)\} = 0$ and $\lambda_{\min}(\bar{M}_k) < 0$. This implies $\lim_{k\to\infty}\{v_k\} = 0$ which implies that for k sufficiently large, $\bar{M}_k$ is positive definite (by virtue of the definition of $\bar{M}_k$), a contradiction. Hence the subsequence $\{\|D_k \mathrm{sgn}(g_k)\| : \lambda_{\min}(\bar{M}_k) < 0\}$ is bounded away from zero and it follows, using Theorem 5, that the corresponding subsequence $\{z_k\}$ is constraint-compatible.

To prove that $\bar{z}_k^T \bar{M}_k \bar{z}_k < \frac{1}{2}\bar{w}_k^T \bar{M}_k \bar{w}_k$ for sufficiently large k, first notice that by Algorithm 7, $\mathcal{S}_k =< D_k^2 \mathrm{sgn}(g_k) >$ only when $\|r(\bar{w}_k, \bar{z}_k)\| < \max(\|D_k g_k\|, -\tau_2 \bar{w}_k^T \bar{M}_k \bar{w}_k)$.

Since $\{D_k g_k\}$ converges to zero and $\lim_{k\to\infty} \lambda_{\min}(\bar{M}_k) < 0$, $\|D_k g_k\| < -\tau_2 \bar{w}_k^T \bar{M}_k \bar{w}_k$ for sufficiently large k, Hence $\|r_k\| = \|r(\bar{w}_k, \bar{z}_k)\| < -\tau_2 \bar{w}_k^T \bar{M}_k \bar{w}_k$.

From

$$r_k = \bar{w}_k - (\bar{z}_k^T \bar{w}_k)\bar{z}_k$$

we have

$$(\bar{z}_k^T \bar{w}_k)^2 \bar{z}_k^T \bar{M}_k \bar{z}_k = \bar{w}_k^T \bar{M}_k \bar{w}_k - 2 r_k^T \bar{M}_k \bar{w}_k + r_k^T \bar{M}_k r_k.$$

But

$$|r_k^T \bar{M}_k \bar{w}_k| \le \rho_M \|r_k\| < \rho_M \tau_2 |\bar{w}_k^T \bar{M}_k \bar{w}_k|,$$

and

$$|r_k^T \bar{M}_k r_k| \le \rho_M \|\bar{r}_k\|^2 < \rho_M^2 \tau_2^2 |\bar{w}_k^T \bar{M}_k \bar{w}_k|,$$

and so

$$(\bar{z}_k^T \bar{w}_k)^2 \bar{z}_k^T \bar{M}_k \bar{z}_k < \bar{w}_k^T \bar{M}_k \bar{w}_k + (2\rho_M \tau_2 + \rho_M^2 \tau_2^2) |\bar{w}_k^T \bar{M}_k \bar{w}_k|.$$

But $\tau_2 < \frac{1}{5\rho_M}$; Therefore,

$$(\bar{z}_k^T \bar{w}_k)^2 \bar{z}_k^T \bar{M}_k \bar{z}_k < \bar{w}_k^T \bar{M}_k \bar{w}_k + \frac{1}{2} |\bar{w}_k^T \bar{M}_k \bar{w}_k| = \frac{1}{2} \bar{w}_k^T \bar{M}_k \bar{w}_k.$$

Finally, since $\bar{z}_k$ and $\bar{w}_k$ are unit vectors, $|\bar{z}_k^T \bar{w}_k| \le 1$; moreover, $\bar{w}_k^T \bar{M}_k \bar{w}_k < 0$ which implies $\bar{z}_k^T \bar{M}_k \bar{z}_k < 0$. Therefore,

$$\bar{z}_k^T \bar{M}_k \bar{z}_k \le \frac{1}{2} \bar{w}_k^T \bar{M}_k \bar{w}_k.$$

■

THEOREM 19. *Assume $\{x_k\}$ is generated by Algorithm 5 with $\{s_k\}$ generated by Algorithm 7 and $\tau_2 < \frac{1}{5\rho_M}$. Then every subsequence $\{s_k\}$ satisfies the consistency condition. Moreover, for any subsequence, if either $\{\|D_k g_k\|\}$ or $\{\max(0, \lambda_{\min}(\bar{M}_k))\}$*

is bounded away from zero, then the corresponding subsequence $\{s_k\}$ is constraint-compatible.

Proof. Applying Theorem 5 to each case in Algorithm 7, it is easy to see that $\{s_k\}$ satisfies consistency.

Assume that if either a subsequence $\{\|D_k g_k\|\}$ or a subsequence $\{\max(0, \lambda_{\min}(\bar{M}_k))\}$ is bounded away from zero. We prove next that the corresponding subsequence $\{s_k\}$ is constraint-compatible.

(i) Suppose there is a subsequence $\{\|D_k g_k\|\}$ bounded from zero. If $\lambda_{\min}(\bar{M}_k) > 0$ then by Algorithm 7 there are three possible ways to compute s_k. All three possibilities clearly yield constraint-compatible sequences $\{s_k\}$ using Theorem 4 and Lemma 17. Assume then that $\lambda_{\min}(\bar{M}_k) \leq 0$. Algorithm 7 gives two possible ways to compute s_k in this case: i.e., $\mathcal{S}_k =< D_k^2 \text{sgn}(g_k) >$ and solve (5.1) to get s_k, or $\mathcal{S}_k =< D_k^2 \text{sgn}(g_k), D_k \bar{w}_k >$ and solve (5.1) to get s_k. In the first case constraint-compatibility of $\{\mathcal{S}_k\}$ follows from the fact that $\{\|D_k \text{sgn}(g_k)\|\}$ is bounded away from zero. In the second case, since $\|r(\bar{w}_k, \bar{z}_k)\| \geq \|D_k g_k\| > 0$, it follows from Lemmas 16 and 17 that $\{s_k\}$ is constraint-compatible.

(ii) Assume $\{D_k g_k\}$ converges to zero, $\lim_{k \to \infty} \lambda_{\min}(\bar{M}_k) < 0$, and $\tau_2 < \frac{1}{5\rho_M}$. Again there are two possible ways in which Algorithm 7 will determine the search direction. Either $\mathcal{S}_k =< D_k^2 \text{sgn}(g_k) >$ and solve (5.1) to get s_k, or $\mathcal{S}_k =< D_k^2 \text{sgn}(g_k), D_k \bar{w}_k >$ and solve (5.1) to get s_k. In the first case constraint-compatibility of $\{\mathcal{S}_k\}$ follows from the fact that $\{\|D_k \text{sgn}(g_k)\|\}$ is bounded from zero. In the second case, since $\|r(\bar{w}_k, \bar{z}_k)\| \geq -\tau_2 \bar{w}_k^T \bar{M}_k \bar{w}_k > 0$, $\{s_k\}$ is constraint-compatible from Lemmas 16 and 17. ∎

The main result follows.

THEOREM 20. *Let $\{x_k\}$ be generated by Algorithm 5 with $\{s_k\}$ generated by Algorithm 7 with $\tau_2 < \frac{1}{5\rho_M}$. Then*

- *Every limit point of $\{x_k\}$ is a first-order point.*
- *Every nondegenerate limit point satisfies the second-order necessary conditions.*
- *If a nondegenerate limit point x_* satisfies second-order sufficiency conditions then, provided τ_g is sufficiently small, $\{x_k\}$ is convergent to x_*; the convergence rate is quadratic, i.e.,*

$$\|x_{k+1} - x_*\| = O(\|x_k - x_*\|^2).$$

Proof. Let $\{s_k\}$ correspond to any subsequence such that either $\{\|D_k g_k\|\}$ or $\{\max(0, \lambda_{\min}(\bar{M}_k))\}$ is bounded away from zero. Then by Theorem 19, the corresponding subsequence $\{s_k\}$ is constraint-compatible. By Theorem 19, $\{s_k\}$ also satisfies the consistency condition. Therefore, by Theorem 14, the result holds for such a subsequence.

Clearly then every subsequence satisfies $\{\|D_k g_k\|\} \to 0$ and $\{\max(0, \lambda_{\min}(\bar{M}_k))\} \to 0$. Hence every limit point of $\{x_k\}$ satisfies first-order and second-order necessary conditions. Let x_* be a limit point satisfying nondegeneracy and second-order sufficiency conditions. By Theorem 13 a unit step size[16] will satisfy (2.1) for $\|x_k - x_*\|$ sufficiently small. Therefore, for $\|x_k - x_*\|$ sufficiently small, Algorithm 5 reduces to Algorithm 6: quadratic convergence follows from Theorem 12. ■

Three computational tasks remain to be discussed before a practical implementable method for the large-scale problem is fully specified. First, the theory demands that $\Delta_k \in [\Delta_l, \Delta_u]$, with $0 < \Delta_l < \Delta_u < \infty$, but imposes no further restriction on Δ_k. In our implementation for minimizing quadratic function subject to bounds, we choose

$$\Delta_k = \min\{\max\{\Delta_l, \|v_k\|\}, \Delta_u\}. \tag{6.1}$$

This choice satisfies the lower and upper bound constraint and is usually commensurate with the distance to the solution, at least with respect to the variables tight at the solution. Experimentally, this choice has performed well.

Second, Algorithm 7 requires that it be determined if $\bar{M}_k$ is positive definite. This can be handled, as we do in our implementation, by attempting a sparse Cholesky factorization (using permutation matrices to limit fill). Iterative methods for sparse linear systems may be possible – this is the subject of ongoing research.

The main computational task yet to be addressed is the determination of a direction $\bar{w}_k$ of sufficient negative curvature [17] such that $\{w_k = D_k \bar{w}_k\}$ also satisfies constraint-compatibility (see Case 2 in Algorithm 7). If a (sparse) Cholesky factorization of $\bar{M}_k$ does not complete then $\bar{M}_k$ is not positive definite and a direction of non-positive curvature, $\bar{w}_k$, is readily available, e.g., [16]. Algorithm 7 can make use of $\bar{w}_k$ provided sufficient negative curvature is displayed by $\bar{w}_k$, i.e.,

$$\bar{w}_k^T \bar{M}_k \bar{w}_k \le \max\{-\epsilon_{nc}, \tau \lambda_{\min}(\bar{M}_k)\}. \tag{6.2}$$

where $\{w_k\}$ is constraint-compatible. A constraint-compatibility test can be designed by introducing a large constant, χ_{cp}, and requiring,

$$\frac{|w_{k_i}|}{|v_{k_i}|} < \chi_{cp}, \quad i = 1:n. \tag{6.3}$$

If either condition (6.2) or condition (6.3) is not satisfied then $\bar{w}_k$ must be rejected. In this case we can turn to a Lanczos process.

Consider that if $\{w_k\}$ is constraint-compatible then $\{D_k \bar{M}_k D_k^{-1} w_k\}$ is also constraint compatible. To see this observe that

$$D_k \bar{M}_k D_k^{-1} w_k = D_k(D_k H_k D_k + J_k^v D_k^{g^+}) D_k^{-1} w_k = (D_k^2 H_k + J_k^v D_k^{g^+}) w_k.$$

[16] If $\alpha_k = 1$ corresponds to a breakpoint then $\alpha_k = \tilde{\alpha}_k = 1 - \epsilon$ where $\tilde{\alpha}_k$ is not a breakpoint, $\tilde{\alpha}_k$ satisfies (2.1), and $\epsilon < \chi_\alpha \|D_k g_k\|$, for some $\chi_\alpha > 0$.

[17] Note: Consistency of $\{w_k = D_k \bar{w}_k\}$ is not an issue. This is because Algorithm 7 uses $\bar{w}_k$ in such a way that consistency of the resulting subsequence $\{s_k\}$ is guaranteed by part 5 of Theorem 5.

Therefore,

$$D_k^{-2}(D_k\bar{M}_kD_k^{-1}w_k) = (H_k + D_k^{-2}J_k^vD_k^{g^+})w_k = (H_kD_k^2 + J_k^vD_k^{g^+})D_k^{-2}w_k.$$

But constraint-compatibility of $\{w_k\}$ means $\{D_k^{-2}w_k\}$ is bounded; by compactness, $\{H_kD_k^2 + J_k^vD_k^{g^+}\}$ is bounded. Therefore, $\{(H_kD_k^2 + J_k^vD_k^{g^+})D_k^{-2}w_k\}$ is bounded, i.e., $\{D_k\bar{M}_kD_k^{-1}w_k\}$ is constraint-compatible.

This argument can be applied recursively: if $\{w_k\}$ is constraint-compatible then $\{w_k^p\}$ is constraint-compatible for any integer m and fixed index k, where

$$w_k^p = (D_k\bar{M}_kD_k^{-1})^m w_k = D_k\bar{M}_k^m D_k^{-1}w_k = D_k\bar{M}_k^m\bar{w}_k. \tag{6.4}$$

Clearly, from (6.4), the Krylov vectors corresponding to matrix $\bar{M}_k$ (and starting vector $\bar{w}_k$), $\bar{w}_k, \bar{M}_k\bar{w}_k, \bar{M}_k^2\bar{w}_k, ...$, yield a set of vectors, $D_k\bar{w}_k$, $D_k\bar{M}_k\bar{w}_k, D_k\bar{M}_k^2\bar{w}_k, ...$, each of which can generate a constraint-compatible sequence provided $\{w_k = D_k\bar{w}_k\}$ is constraint-compatible.

So the Krylov vectors, with matrix $\bar{M}_k$ and starting vector $\bar{w}_k$, generate constraint-compatible sequences. Let $\mathcal{K}_k(m_k, \bar{w}_k)$ be the Krylov space generated by the first m_k Krylov vectors, $\bar{w}_k, \bar{M}_k\bar{w}_k, \bar{M}_k^2\bar{w}_k, ..., \bar{M}_k^{m_k-1}\bar{w}_k$ for some vector $\bar{w}_k$ where $\{w_k = D_k\bar{w}_k\}$ is constraint-compatible.

An interesting and important question is this: Does a sequence of Krylov subspaces $\{\mathcal{K}_k\}$ generate a sequence of matrix products $\{D_kY_k\}$ satisfying the conditions of Lemma 16 where the columns of Y_k are orthonormal and $< Y_k >= \mathcal{K}_k$? The answer is yes provided the Krylov vectors defining subspace $\mathcal{K}_k$ are sufficiently linearly independent for every k.

THEOREM 21. *Assume $\{w_k\}$ is a constraint-compatible sequence and define $\bar{w}_k = D_k^{-1}w_k$. Let $\mathcal{K}_k = \mathcal{K}_k(m_k, \bar{w}_k\}$ be the Krylov space defined by the Krylov vectors $\bar{w}_k, \bar{M}_k\bar{w}_k, \bar{M}_k^2\bar{w}_k, ..., \bar{M}_k^{m_k-1}\bar{w}_k$. Further, assume that $|\mu_k| > \tau > 0, \forall k$, where μ_k is the subdiagonal of the tridiagonal matrix $T_k = Y_k^T\bar{M}_kY_k = diag(\lambda_k, 0) + diag(\mu_k, 1) + diag(\mu_k, -1)$ obtained [18] from the Lanczos method with $Y_k^TY_k = I$. Then each column of D_kY_k generates a constraint-compatible sequence.*

Proof. Assume that $Y_k = [y_k^1, \cdots, y_k^{m_k}]$. Note that $\mu_k > \tau$ implies that $m_k \le n$. The Lanczos vectors $\{y_k^i\}$ satisfy $\bar{M}_ky_k^1 = \lambda_k^1y_k^1 + \mu_k^2y_k^2$ and for $1 \le i \le m_k - 1$ (see [19], page 477),

$$\bar{M}_ky_k^i = \mu_k^{i-1}y_k^{i-1} + \lambda_k^iy_k^i + \mu_k^{i+1}y_k^{i+1}, \tag{6.5}$$

where $\mu_k^0y_k^0 \stackrel{\text{def}}{=} 0$. Moreover, for $1 \le i \le m_k$,

$$\lambda_k^i = (y_k^i)^T\bar{M}_ky_k^i, \quad \mu_k^i = \|r_k^i\|, \tag{6.6}$$

[18] The matrix diag$(\lambda_k, 0)$ denotes a diagonal matrix with the diagonal defined by vector λ_k; matrix diag$(\mu_k, 1)$ is a zero matrix except for the main super-diagonal which is defined by vector μ_k; matrix diag$(\mu_k, -1)$ is the zero matrix except for the main sub-diagonal which is defined by vector μ_k

where $r_k^1 = (\bar{M}_k - \lambda_k^1 I)y_k^1$ and for $2 \le i \le m_k$,

$$r_k^i = (\bar{M}_k - \lambda_k^i I)y_k^i - \mu_k^{i-1} y_k^{i-1}. \tag{6.7}$$

Following the usual Lanczos procedure, $y_k^1 = \frac{\bar{w}_k}{\|\bar{w}_k\|}$, and so by assumption of constraint-compatibility of $\{w_k\}$, $\{D_k y_k^1\}$ is constraint-compatible. Clearly, from (6.6), the boundedness of $\{\bar{M}_k\}$, and the orthonormality of Y_k, for $1 \le i \le m_k$ the sequence $\{\lambda_k^i\}$ is bounded.

From $r_k^1 = (\bar{M}_k - \lambda_k^1 I)y_k^1$ and the boundedness of $\{\bar{M}_k\}$ and $\{\lambda_k^i\}$, $\{\mu_k^1\}$ is bounded. By a simple induction on i and (6.6), we conclude that $\{\mu_k^i\}$, $1 \le i \le m$, is also bounded. Using the assumption that $|\mu_k| > \tau > 0$, (6.5) and a simple induction on i, $\{D_k y_k^i\}$ is constraint-compatible for $1 \le i \le m_k$. ■

Theorem 21 tells us that the usual Lanczos procedure will produce an orthonormal basis Y_k of the Krylov subspace $\mathcal{K}_k$ such that each column of $D_k Y_k$ generates a constraint-compatible sequence, provided the main subdiagonal elements of T_k are bounded away from zero. Fortunately, as discussed in [12], page 139, it is quite reasonable to assume that until all of the distinct eigenvalues of the original matrix have been approximated well by eigenvalues of the Lanczos matrices, all of the off-diagonal entries are uniformly bounded away from zero, i.e., $\mu_i \ge \tau_\mu, 1 \le i \le j$ for some $\tau_\mu > 0$. Therefore, the Lanczos procedure can be continued until an eigenvector of T_k is found, say $\hat{w}_k$, such that (6.2) is satisfied, i.e.,

$$\bar{w}_k^T \bar{M}_k \bar{w}_k \le \max\{-\epsilon_{nc}, \tau\lambda_{\min}(\bar{M}_k)\},$$

where $\|\bar{w}_k\|_2 = 1$ and $0 < \tau < 1$. But since every column of $D_k Y_k$ generates a constraint-compatible sequence and $\|\bar{w}_k\|_2 = 1$, $\{w_k = D_k \bar{w}_k\}$ is constraint-compatible. Therefore, $\bar{w}_k$ can be used to satisfy both (6.2) and (6.3).

A good starting vector for the Lanczos procedure is $w_k = D_k^2 \mathrm{sgn}(g_k)$. This choice yields a constraint-compatible sequence $\{w_k\}$ and is bounded from zero (except when x_* is a vertex in which case the need for a Lanczos procedure does not arise).

7. Concluding Remarks. We have proposed a new method, a reflective Newton method, for solving nonlinear minimization problems where some of the variables have upper and/or lower bounds. We have established strong convergence properties. In particular, reflective Newton methods can achieve global and quadratic convergence.

The proposed reflective Newton method involves the solution of a reduced trust region problem, (5.1). In (5.1), subspace $\mathcal{S}_k$ must be chosen with extreme care to ensure the second-order convergence properties and to maintain practical viability in the large-scale setting. In this paper we show that a small dimensional subspace can be used, i.e., $|\mathcal{S}_k| \le 2$, and yet the attractive convergence properties obtained with $\mathcal{S}_k = \Re^n$ can be maintained. Our method involves the use of a sparse Cholesky factorization as well as a Lanczos procedure used to construct $\mathcal{S}_k$.

Experimental results for the case when the objective function is quadratic are provided in [10]. These computational results are extremely encouraging and indicate that reflective Newton methods have strong potential for large-scale computations. Experimentation on general nonlinear functions is a current research activity and results will be available in a future report.

Research on two extensions of this work is underway. First, we are studying *inexact* reflective Newton methods for problem (1.1). Our current implementation rests on a (partial) sparse Cholesky factorization of $\bar{M}_k$. A limitation with this approach is that a (partial) sparse Cholesky factorization is not always economical. Therefore, we are considering a reflective Newton procedure that only requires the iterative use of $\bar{M}_k$.

Second, we are studying the adaptation of reflective Newton methods to bound-constrained problems with additional linear equality constraints:

$$\min_x \{f(x): \; Ax = b, \; l \le x \le u\}. \tag{7.1}$$

If we assume then that x_k is a feasible point then, following the lines in this paper, a feasible descent direction can be obtained by solving

$$\min_s \{s^T g_k + \frac{1}{2} s^T M_k s : \; \|D_k^{-1} s\|_2 \le \Delta_k, s \in \mathcal{S}_k\} \tag{7.2}$$

where $M(x) = H + J_v D^{\frac{g}{v}}$, and $\mathcal{S}_k$ is contained in the null space of matrix A. We have already sketched a technique in this paper for solving such problems; however, this approach may not be practical here (in general) since in this case $|\mathcal{S}_k|$ is not necessarily small. Therefore, a different sparsity-preserving method must be used to solve (7.2) – Coleman and Hempel [4] have developed a technique based on the use of an "augmented" system that may have some potential here.

A possible reflective Newton approach to problem (7.1) is clear from a geometric point of view. After generating a search direction from a strictly feasible point x_k, using (7.2), a piecewise linear (reflective) path can be searched to find a new (improved) point. Nevertheless, despite this clear geometric picture, many research issues remain, not the least of which is the efficient calculation of this piecewise linear path (while exploiting and maintaining sparsity).

8. Acknowledgements. We thank our colleague Jianguo Liu for many helpful remarks on this work.

REFERENCES

[1] L. Armijo, *Minimization of functions having lipschitz continuous first partial derivatives*, Pac. J. Math., 16 (1966), pp. 1–3.

[2] Å. Björck, *A direct method for sparse least squares problems with lower and upper bounds*, Numerische Mathematik, 54 (1988), pp. 19–32.

[3] R. H. Byrd and R. B. Schnabel, *Approximate solution of the trust region problem by minimization over two-dimensional subspaces*, Mathematical Programming, 40 (1988), pp. 247–263.

[4] T. F. Coleman and C. Hempel, *Computing a trust region step for a penalty function*, SIAM Journal on Scientific and Statistical Computing, 11 (1990), pp. 180–201.

[5] 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.

[6] ———, *A globally and superlinearly convergent algorithm for convex quadratic programs with simple bounds*, Tech. Rep. TR 90-1092, Computer Science Department, Cornell University, February, 1990 (to appear in SIAM Journal on Optimization).

[7] T. F. Coleman and Y. Li, *A quadratically-convergent algorithm for the linear programming problem with lower and upper bounds*, in Large-Scale Numerical Optimization, T. F. Coleman and Y. Li, eds., SIAM, 1990, pp. 49–57. Proceedings of the Mathematical Sciences Institute workshop, October 1989, Cornell University.

[8] ———, *A global and quadratically-convergent method for linear l_∞ problems*, SIAM Journal on Numerical Analysis, 29 (1992), pp. 1166–1186.

[9] ———, *A globally and quadratically convergent affine scaling method for linear l_1 problems*, Mathematical Programming, 56, Series A (1992), pp. 189–222.

[10] ———, *A reflective Newton method for minimizing a quadratic function subject to bounds on the variables*, Tech. Rep. TR 92-1315, Computer Science Department, Cornell University, 1992.

[11] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Testing a class of methods for solving minimization problems with simple bounds on the variables*, Mathematics of Computation, 50 (1988), pp. 399–430.

[12] J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large symmetric eigenvalue computations, Vol. 1 Theory*, Birkha'user, Boston, 1985.

[13] R. S. Dembo and U. Tulowitzki, *On the minimization of quadratic functions subject to box constraints*, Tech. Rep. B 71, Yale University, 1983.

[14] R. Fletcher and M. P. Jackson, *Minimization of a quadratic function of many variables subject only to lower and upper bounds*, Journal of the Institute for Mathematics and its Applications, 14 (1974), pp. 159–174.

[15] P. Gill and W. Murray, *Minimization subject to bounds on the variables*, Tech. Rep. Report NAC 71, National Physical Laboratory, England, 1976.

[16] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, 1981.

[17] D. Goldfarb, *Curvilinear path steplength algorithms for minimization algorithms which use directions of negative curvature*, Mathematical Programming, 18 (1980), pp. 31–40.

[18] A. Goldstein, *On steepest descent*, SIAM Journal on Control, 3 (1965), pp. 147–151.

[19] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, 1989.

[20] J. J. Júdice and F. M. Pires, *Direct methods for convex quadratic programs subject to box constraints*, departamento de matemática, Universidade de Coimbra, 3000 Coimbra, Portugal, 1989.

[21] Y. Li, *A globally convergent method for l_p problems*, Tech. Rep. 91-1212, Computer Science Dept., Cornell University, 1991 (to appear in SIAM Journal on Optimization).

[22] P. Lotstedt, *Solving the minimal least squares problem subject to bounds on the variables*, BIT, 24 (1984), pp. 206–224.

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

[24] D. P. O'Leary, *A generalized conjugate gradient algorithm for solving a class of quadratic programming problems*, Linear Algebra and its Applications, 34 (1980), pp. 371–399.

[25] U. Öreborn, *A direct method for sparse nonnegative least squares problems*, PhD thesis, Department of Mathematics, Linköping University, Linköping, Sweden, 1986.

[26] J. M. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970.

[27] 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(1) (1985), pp. 47–67.

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

[29] E. K. Yang and J. W. Tolle, *A class of methods for solving large convex quadratic programs subject to box constraints*, tech. rep., Department of Operations Research, University of North Carolina, Chapel Hill, North Carolina, 1988.