Lectures in Applied Mathematics
Volume 26 (1990)

On Characterizations of Superlinear Convergence for Constrained Optimization

THOMAS F. COLEMAN

Abstract. We show how the Dennis–Moré characterization of superlinear convergence for unconstrained optimization can be applied, and usefully restricted, for use in the constrained setting.

1. Introduction. The main purpose of this paper is to illustrate how the Dennis–Moré [7] characterization of superlinear convergence can be adapted to the (equality) constrained optimization setting. In particular, we follow Goodman [11] and replace the constrained minimization problem with a smooth zero-finding problem (valid in a neighborhood of the solution). It is then possible to apply the Dennis–Moré characterization directly. However, there is a subtle point: the function in the zero-finding problem is smooth but is not computable because its definition depends on information computable at the solution only. Nevertheless, we demonstrate the usefulness of this viewpoint.

In Section 2 we first consider the case where exact second derivatives are known; we provide a new and easy proof of quadratic convergence of algorithms in the sequential quadratic programming (SQP)

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 65H10, 65K05, 65K10.

Key words and phrases. Constrained optimization, equality constraints, numerical optimization, quasi-Newton method, secant method, superlinear characterization, sequential quadratic programming, SQP-method.

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 the U. S. Army Research Office through the Mathematical Sciences Institute, Cornell University.

class. Section 3 follows with a discussion of characterizations of superlinear convergence and various restrictions. The Boggs–Tolle–Wang [1] "characterization" is shown to be a restriction of the general characterization (i.e., it is applicable to a subclass of methods only). Finally, in Section 4, we provide a new proof of the superlinear convergence of the Coleman–Conn [4] method, using the results developed in the previous section.

2. Quadratic convergence of sequential quadratic programming (SQP). We are interested in the nonlinear equality constrained problem,

$$\text{minimize } \{f(\mathbf{x}) : \mathbf{c}(\mathbf{x}) = \mathbf{0}\}$$

where $f: \mathbf{R}^n \to \mathbf{R}$, $\mathbf{c}: \mathbf{R}^n \to \mathbf{R}^t$, $t \le n$, and $\mathbf{c}(\mathbf{x}) = (c_1(\mathbf{x}), \ldots, c_t(\mathbf{x}))^{\mathrm{T}}$. In this section we provide a simple proof of the quadratic convergence of the SQP method for problem (1). The proof is new though it is inspired by the viewpoint developed by Goodman [11]; however, in our approach we use a smooth zero-finding problem valid in a neighborhood of the solution, whereas Goodman defines a sequence of smooth zero-finding problems, each valid around the current point.

The technique and terminology introduced in this section will be useful in the remainder of the paper. Let $\mathbf{A}(\mathbf{x})$ denote the matrix of constraint gradients $\mathbf{A}(\mathbf{x}) = (\nabla c_1(\mathbf{x}), \ldots, \nabla c_t(\mathbf{x}))$. The following assumptions are frequently used in this paper:

(A_1) f and $c_i, i = 1, t$, are twice continuously differentiable on an open convex set D;

(A_2) $\mathbf{x}_* \in D$, where x_* is a local solution to (1); $\mathbf{A}_* \overset{\text{def}}{=} \mathbf{A}(\mathbf{x}_*)$ is of full column rank t;

(A_3) $\nabla^2 f_* + \sum \lambda_i^* \nabla^2 c_i^*$ is positive definite on null($\mathbf{A}_*^{\mathrm{T}}$), where $\nabla f_* + \mathbf{A}_* \lambda_* = 0$;

(A_4) $\lambda(\mathbf{x})$ is a continuous function on D satisfying a Lipschitz condition at $\mathbf{x}_*$.

We say that $\{\mathbf{x}_k\}$ is generated by the SQP method if $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k^{\mathrm{SQP}}$ and $\mathbf{x}_k^{\mathrm{SQP}}$ solves

$$\min \left\{ \mathbf{s}^{\mathrm{T}} \nabla L_k + \tfrac{1}{2} \mathbf{s}^{\mathrm{T}} \nabla^2 L_k \mathbf{s} : \mathbf{A}_k^{\mathrm{T}} \mathbf{s} = -\mathbf{c}_k \right\}, \tag{2}$$

where $L_k(\mathbf{x}) = f(\mathbf{x}) + \lambda_k^{\mathrm{T}} \mathbf{c}(\mathbf{x})$, $\lambda_k = \lambda(\mathbf{x}_k)$, and λ satisfies A_4. (For example, the least-squares multiplier function is given by $\lambda(\mathbf{x}) = -\mathbf{R}(\mathbf{x})^{-1} \mathbf{Y}(\mathbf{x})^{\mathrm{T}} \nabla f(\mathbf{x})$ where $\mathbf{A}(\mathbf{x}) = \mathbf{Y}(\mathbf{x}) \mathbf{R}(\mathbf{x})$; $\mathbf{Y}$ is n-by-t with orthonormal columns, $\mathbf{R}$ is t-by-t and upper triangular.)

Notational note. For simplicity we will write L_k (i.e., without argument) to mean $L_k(\mathbf{x}_k)$. In addition, we write ∇L_k and $\nabla^2 L_k$ to refer to the gradient and Hessian of $L_k(\mathbf{x}_k)$, respectively, with λ_k treated as a constant.

If $\mathbf{A}_k$ is of full column rank t and the two-sided projection of $\nabla^2 L_k$ onto null$(\mathbf{A}_k^{\mathrm{T}})$ is positive definite, then (2) has a unique solution. Specifically, let $\mathbf{Z}_k$ be *any basis* for null$(\mathbf{A}_k^{\mathrm{T}})$ ($\mathbf{Z}_k$ is not necessarily an orthogonal basis). We can write $\mathbf{s} = \mathbf{Y}_k\mathbf{v}_k + \mathbf{Z}_k\mathbf{h}$, where $\mathbf{v}_k = -\mathbf{R}_k^{-\mathrm{T}}\mathbf{c}_k$. Hence, (2) is equivalent to

$$\text{(3)} \qquad \text{minimize}\, \mathbf{h}^{\mathrm{T}}\mathbf{Z}_k^{\mathrm{T}}(\nabla f_k + \nabla^2 L_k \mathbf{Y}_k\mathbf{v}_k) + \tfrac{1}{2}h^{\mathrm{T}}\mathbf{Z}_k^{\mathrm{T}}\nabla^2 L_k \mathbf{Z}_k\mathbf{h}.$$

But we have assumed that $\mathbf{Z}_k^{\mathrm{T}}\nabla^2 L_k\mathbf{Z}_k > 0$ and therefore the solution to (3) is given by

$$\text{(4)} \qquad (\mathbf{Z}_k^{\mathrm{T}}\nabla^2 L_k\mathbf{Z}_k)\mathbf{h}_k = -\mathbf{Z}_k^{\mathrm{T}}(\nabla f_k + \nabla^2 L_k\mathbf{Y}_k\mathbf{v}_k)$$

and the solution to (2) is

$$\text{(5)} \qquad \mathbf{s}_k^{\mathrm{SQP}} = \mathbf{Z}_k\mathbf{h}_k + \mathbf{Y}_k\mathbf{v}_k.$$

But (4) and (5) can be combined to give

$$\begin{pmatrix} \mathbf{Z}_k^{\mathrm{T}}\nabla^2 L_k \\ \mathbf{A}_k^{\mathrm{T}} \end{pmatrix} (\mathbf{Z}_k\mathbf{h}_k + \mathbf{Y}_k\mathbf{v}_k) = \begin{pmatrix} -\mathbf{Z}_k^{\mathrm{T}}\nabla f_k \\ -\mathbf{c}_k \end{pmatrix}$$

which is equivalent to

$$\text{(6)} \qquad \begin{pmatrix} \mathbf{Z}_k^{\mathrm{T}}\nabla^2 L_k \\ A_k^{\mathrm{T}} \end{pmatrix} \mathbf{s}_k^{\mathrm{SQP}} = \begin{pmatrix} -\mathbf{Z}_k^{\mathrm{T}}\nabla f_k \\ -\mathbf{c}_k \end{pmatrix}.$$

Note that $\mathbf{s}_k^{\mathrm{SQP}}$ is independent of the particular basis $\mathbf{Z}_k$.

To be more precise, we should include a dependence on λ when we refer to an SQP direction, e.g., $\mathbf{s}_k^{\mathrm{SQP}_\lambda}$, because the SQP direction will differ with different λ-rules. However, the differences are low-order: All results presented in this section are valid for any continuous λ-function satisfying a Lipschitz condition at $\mathbf{x}_*$, and therefore we suppress the λ subscript. We do assume that the rule for choosing λ is consistently applied so that $\lambda(\mathbf{x})$ can be viewed as a specific continuous function, Lipschitz continuous at $\mathbf{x}_*$.

In order to analyze the convergence behavior of $\{\mathbf{x}_k\}$, where $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k^{\mathrm{SQP}}$, we need the following result (Theorem 3.4 of [8]).

LEMMA 1. *Let* $\mathbf{F}: \mathbf{R}^n \to \mathbf{R}^n$ *satisfy*

(a) $\mathbf{F}$ *is continuously differentiable on an open convex set* D;

(b) *there exists* $\mathbf{x}_* \in D$ *such that* $\mathbf{F}(\mathbf{x}_*) = 0$ *and* $\mathbf{F}'(\mathbf{x}_*)$ *is nonsingular*;

(c) $\|\mathbf{F}'(\mathbf{x}) - \mathbf{F}'(\mathbf{x}_*)\| \le \eta(\|\mathbf{x} - \mathbf{x}_*\|)$ *for all* $\mathbf{x} \in D$ *and some* $\eta \ge 0$.

Assume $\{\mathbf{x}_k\} \in D$ *and* $\{\mathbf{x}_k\} \to \mathbf{x}_*$. *If there exists a sequence of nonsingular matrices* $\{\mathbf{B}_k\}$ *such that* $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$, $\mathbf{B}_k\mathbf{s}_k = -\mathbf{F}(\mathbf{x}_k)$, *and* $\|\mathbf{B}_k - \mathbf{F}'_*\| = O(\|\mathbf{x}_k - \mathbf{x}_*\|)$ *then* $\{\mathbf{x}_k\}$ *converges quadratically to* $\mathbf{x}_*$.[1]

In order to apply Lemma 1 we must phrase (1) as a continuously differentiable system of equations (at least in a neighborhood of a solution $\mathbf{x}_*$). System (6) provides a strong hint as to a possible form; however, a difficulty arises because the standard method of computing $\mathbf{Z}(\mathbf{x})$ does not yield a continuously differentiable basis representation [5]. Moreover, Byrd and Schnabel [3] have shown that a smooth representation of a basis of the null space does not exist, in general. Nevertheless, various (theoretical) forms can be used [3], [11] to yield a continuously differentiable representation in the neighborhood of any given point. Below we derive and use a specific form; other forms are possible. Indeed, it is possible to proceed without using an explicit form; however, in the interest of clarity, we prefer to be concrete. Specifically, let $\bar{\mathbf{x}} \in \mathbf{R}^n$ and suppose that $\bar{\mathbf{A}} \overset{\text{def}}{=} \mathbf{A}(\bar{\mathbf{x}})$ is of full rank t. Let $\bar{\mathbf{Z}}$ be any basis for null$(\bar{\mathbf{A}}^{\mathrm{T}})$ and define

$$\mathbf{Z}(\mathbf{x}) = (\mathbf{I} - \bar{\mathbf{A}}(\mathbf{A}^{\mathrm{T}}\bar{\mathbf{A}})^{-1}\mathbf{A}^{\mathrm{T}})\bar{\mathbf{Z}}. \tag{7}$$

In a neighborhood of $\bar{\mathbf{x}}$, $\mathbf{Z}(\mathbf{x})$ is continuously differentiable with

$$\dot{\mathbf{Z}} = \bar{\mathbf{A}}(\mathbf{A}^{\mathrm{T}}\bar{\mathbf{A}})^{-1}\dot{\mathbf{A}}^{\mathrm{T}}\bar{\mathbf{A}}(\mathbf{A}^{\mathrm{T}}\bar{\mathbf{A}})^{-1}\mathbf{A}^{\mathrm{T}}\bar{\mathbf{Z}} - \bar{\mathbf{A}}(\mathbf{A}^{\mathrm{T}}\bar{\mathbf{A}})^{-1}\dot{\mathbf{A}}^{\mathrm{T}}\bar{\mathbf{Z}}. \tag{8}$$

Therefore,

$$\dot{\mathbf{Z}}(\bar{\mathbf{x}}) = -\bar{\mathbf{A}}(\bar{\mathbf{A}}^{\mathrm{T}}\bar{\mathbf{A}})^{-1}\dot{\mathbf{A}}(\bar{\mathbf{x}})^{\mathrm{T}}\bar{\mathbf{Z}} = -\bar{\mathbf{Y}}\bar{\mathbf{R}}^{-\mathrm{T}}\dot{\mathbf{A}}(\bar{\mathbf{x}})\bar{\mathbf{Z}}, \tag{9}$$

where $\bar{\mathbf{A}} = \bar{\mathbf{Y}}\bar{\mathbf{R}}$. (Note that $\dot{\mathbf{Z}}(\bar{\mathbf{x}})^{\mathrm{T}}\nabla f(\bar{\mathbf{x}}) = \bar{\mathbf{Z}}^{\mathrm{T}}\sum\bar{\lambda}_i\nabla^2 c_i(\bar{\mathbf{x}})$, where $\bar{\lambda} = -\bar{\mathbf{R}}^{-1}\bar{\mathbf{Y}}^{\mathrm{T}}\nabla f(\bar{\mathbf{x}})$.)

Using this definition of $\mathbf{Z}$ defined around $\bar{\mathbf{x}} = \mathbf{x}_*$, consider the nonlinear system $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, where

$$\mathbf{F}(\mathbf{x}) \overset{\text{def}}{=} \begin{pmatrix} \mathbf{Z}(\mathbf{x})^{\mathrm{T}}\nabla f(\mathbf{x}) \\ \mathbf{c}(\mathbf{x}) \end{pmatrix}. \tag{10}$$

Clearly, under assumptions (A_1)–(A_3), $\mathbf{x}_*$ is an isolated zero of (10).

Note that

$$\mathbf{J}(\mathbf{x}) \overset{\text{def}}{=} \mathbf{F}'(\mathbf{x}) = \begin{pmatrix} \dot{\mathbf{Z}}(\mathbf{x})^{\mathrm{T}}\nabla f(\mathbf{x}) + \mathbf{Z}(\mathbf{x})^{\mathrm{T}}\nabla^2 f(\mathbf{x}) \\ \mathbf{A}^{\mathrm{T}}(\mathbf{x}) \end{pmatrix} \tag{11}$$

[1] We make extensive use of the "O" and "o" notation: $\varphi_k = O(\psi_k)$ means that the ratio φ_k/ψ_k remains bounded as $k \to \infty$ and $\varphi_k = o(\psi_k)$ means that the ratio $\varphi_k/\psi_k \to 0$ as $k \to \infty$.

and hence

$$(12)\qquad \mathbf{J}_* = \begin{pmatrix} \mathbf{Z}_*^{\mathrm{T}}(\nabla^2 f_* + \sum \lambda_i^* \nabla^2 c_i^*) \\ \mathbf{A}_*^{\mathrm{T}} \end{pmatrix}.$$

Assumptions (A_2) and (A_3) imply that $\mathbf{J}_*$ is nonsingular.

Unfortunately, the SQP step is not a Newton step for (10) and therefore quadratic convergence is not automatic. Indeed, the construction of a Newton process for (10) is unclear since

$$(13)\qquad \mathbf{Z}(\mathbf{x}) = (\mathbf{I} - \mathbf{A}_*(\mathbf{A}(\mathbf{x})^{\mathrm{T}}\mathbf{A}_*)^{-1}\mathbf{A}(\mathbf{x})^{\mathrm{T}})\mathbf{Z}_*$$

and typically $\mathbf{Z}_*$ and $\mathbf{A}_*$ are unknown, except at $\mathbf{x}_*$.

Our approach to this problem is to show that the solution to (6) (which is independent of the choice of $\mathbf{Z}_k$) is actually a solution to an *approximate* Newton system, $\mathbf{B}_k \mathbf{s}_k^{\mathrm{SQP}} = -\mathbf{F}(\mathbf{x}_k)$, where $\|\mathbf{B}_k - \mathbf{J}_*\| = O(\|\mathbf{x}_k - \mathbf{x}_*\|)$. In particular, since $\mathbf{s}_k^{\mathrm{SQP}}$ is independent of the choice of basis $\mathbf{Z}_k$ in (6), $\mathbf{s}_k^{\mathrm{SQP}}$ satisfies

$$(14)\qquad \begin{pmatrix} \mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla^2 L_k \\ \mathbf{A}_k^{\mathrm{T}} \end{pmatrix} \mathbf{s}_k^{\mathrm{SQP}} = \begin{pmatrix} -\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla f_k \\ -c_k \end{pmatrix}$$

where $\mathbf{Z}(\mathbf{x}_k)$ is defined by (13). Define

$$(15)\qquad \mathbf{B}_k = \begin{pmatrix} \mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla^2 L_k \\ \mathbf{A}_k^{\mathrm{T}} \end{pmatrix}.$$

Hence, $\mathbf{s}_k^{\mathrm{SQP}}$ satisfies $\mathbf{B}_k \mathbf{s}_k^{\mathrm{SQP}} = -\mathbf{F}(\mathbf{x}_k)$.

LEMMA 2. *Let assumptions* (A_1)–(A_4) *hold. Further, assume that* $\nabla^2 f$, $\nabla^2 c_i$ $(i = 1, \ldots, t)$ *are Lipschitz continuous at* $\mathbf{x}_*$. *Let* $\mathbf{x}_c$ *be an arbitrary point and define* $L_c(\mathbf{x}) = f(\mathbf{x}) + \lambda_c^{\mathrm{T}}\mathbf{c}(\mathbf{x})$, *where* $\lambda_c = \lambda(\mathbf{x}_c)$; *define*

$$\mathbf{B}_c(\mathbf{x}) = \begin{pmatrix} \mathbf{Z}(\mathbf{x})^{\mathrm{T}}\nabla^2 L_c(\mathbf{x}) \\ \mathbf{A}(\mathbf{x})^{\mathrm{T}} \end{pmatrix}.$$

Then, for all $\mathbf{x}_c$ *sufficiently close to* $\mathbf{x}_*$, $\|\mathbf{B}_c(\mathbf{x}_c) - \mathbf{J}_*\| = O(\|\mathbf{x}_c - \mathbf{x}_*\|)$.

PROOF. First consider rows $1, \ldots, n-t$:

$$\begin{aligned}(\mathbf{I}_{n-t}, \mathbf{0})(\mathbf{B}_c(\mathbf{x}_c) - \mathbf{J}_*) &= \mathbf{Z}(\mathbf{x}_c)^{\mathrm{T}}\nabla^2 L_c(\mathbf{x}_c) - \mathbf{Z}_*^{\mathrm{T}}(\nabla^2 L_*(\mathbf{x}_*)) \\ &= (\mathbf{Z}(\mathbf{x}_c)^{\mathrm{T}} - \mathbf{Z}_*^{\mathrm{T}})(\nabla^2 L_*(\mathbf{x}_*)) \\ &\quad + \mathbf{Z}(\mathbf{x}_c)^{\mathrm{T}}(\nabla^2 L_c(\mathbf{x}_c) - \nabla^2 L_*(\mathbf{x}_*)).\end{aligned}$$

But $\dot{\mathbf{Z}}$ is bounded in a neighborhood of $\mathbf{x}_*$ and therefore, using Taylor's theorem, we have

$$(16)\qquad \|\mathbf{Z}(\mathbf{x}_c) - \mathbf{Z}_*\| = O(\|\mathbf{x}_c - \mathbf{x}_*\|)$$

for $\mathbf{x}_c$ in a neighborhood of $\mathbf{x}_*$; using the Lipschitz continuity of $\nabla^2 f$, $\nabla^2 c_i$, and λ, it follows that

$$(17) \qquad ||\nabla^2 L_c(\mathbf{x}_c) - \nabla^2 L_*(\mathbf{x}_*)|| = O(||\mathbf{x}_c - \mathbf{x}_*||)$$

in a neighborhood of $\mathbf{x}_*$. Equations (16) and (17) yield

$$(18) \qquad ||(\mathbf{I}_{n-t}, \mathbf{0})(\mathbf{B}_c - \mathbf{J}_*)|| = O(||\mathbf{x}_c - \mathbf{x}_*||),$$

for all $\mathbf{x}_c$ in a neighborhood of $\mathbf{x}_*$. Next consider the last t rows:

$$(\mathbf{0}, \mathbf{I}_t)(\mathbf{B}_c - \mathbf{J}_*) = (\mathbf{A}_c^{\mathrm{T}} - \mathbf{A}_*^{\mathrm{T}}).$$

But $\dot{\mathbf{A}}(\mathbf{x})$ is bounded in a neighborhood of $\mathbf{x}_*$ and so by Taylor's theorem we get

$$(19) \qquad ||\mathbf{A}_c - \mathbf{A}_*|| = O(||\mathbf{x}_c - \mathbf{x}_*||)$$

for all $\mathbf{x}_c$ sufficiently close to $\mathbf{x}_*$.

The result now follows from (18) and (19). ■

We are now poised to use Lemma 1 to prove quadratic convergence of the SQP method; we need only establish assumption (c) of Lemma 1 with respect to our definition of $\mathbf{F}$. However, the fact that $\dot{\mathbf{A}}(\mathbf{x})$ is bounded in a neighborhood of $\mathbf{x}_*$ and definition (8) yield

$$(20) \qquad ||\dot{\mathbf{Z}}(\mathbf{x}) - \dot{\mathbf{Z}}_*|| = O(||\mathbf{x} - \mathbf{x}_*||)$$

for all $\mathbf{x}$ sufficiently close to $\mathbf{x}_*$. Next, using (20), the Lipschitz continuity of $\nabla^2 f$, and the boundedness of $\dot{\mathbf{A}}(\mathbf{x})$, we obtain the required result:

$$(21) \qquad ||\mathbf{J}(\mathbf{x}) - \mathbf{J}_*|| = O(||\mathbf{x} - \mathbf{x}_*||)$$

for all $\mathbf{x}$ sufficiently close to $\mathbf{x}_*$.

We now have all the necessary prerequisites to establish quadratic convergence of the SQP method.

THEOREM 3. *Let assumptions* (A_1)–(A_4) *hold; assume that* $\nabla^2 f$, $\nabla^2 c_i$, $i = 1, \ldots, t$, *are Lipschitz continuous at* $\mathbf{x}_*$. *Assume* $\{\mathbf{x}_k\} \in D$ *and* $\{\mathbf{x}_k\} \to \mathbf{x}_*$, *where* $\{\mathbf{x}_k\}$ *is generated by the* SQP *method. Then* $\{\mathbf{x}_k\}$ *converges quadratically.*

PROOF. Clearly, assumptions (a), (b), and (c) of Lemma 1 hold. Moreover, by Lemma 2 there exists a $\mathbf{B}_k$ such that $\mathbf{B}_k \mathbf{s}_k^{\mathrm{SQP}} = -\mathbf{F}(\mathbf{x}_k)$ and $||\mathbf{B}_k - \mathbf{J}_*|| = O(||\mathbf{x}_k - \mathbf{x}_*||)$, provided D is small enough. The result follows.

The local convergence of the SQP method is easily established by Theorem 5.1 in [8].

REMARK. Goodman's [11] proof of quadratic convergence differs from ours in the following respect. Goodman's technique involves the fact that the SQP steps—with the least-squares multipliers—form a sequence of *exact* Newton steps for a *sequence* of functions, $\mathbf{F}_i$. In contrast, we use the fact that the SQP step—with any Lipschitz continuous multiplier function—is an *approximate* Newton step for a *fixed* function $\mathbf{F}$. Our proof is also closely related to the approach in Tapia [14] in which an approximate Newton point of view is also taken. Tapia's development differs in that a projection-based function is used instead of defining a smooth basis $\mathbf{Z}$.

3. Superlinear convergence for constrained optimization algorithms. For unconstrained minimization the characterization of Dennis and Moré ([7], Theorem 2.2 or [8], Theorem 3.1) has proven to be very useful. It can also be used for constrained optimization, using $\mathbf{F}(\mathbf{x})$ and $\mathbf{J}_*$ defined in (10) and (12), respectively. For completeness we reproduce the result here, reworded to reflect the constrained optimization setting.

THEOREM 4. *Let assumptions* (A_1)–(A_3) *hold. Let* $\{\mathbf{M}_k\}$ *be a sequence of nonsingular matrices. Define* $\mathbf{F}$ *and* $\mathbf{J}_*$ *by* (7), (10), *and* (12). *Suppose that for some* $\mathbf{x}_0$ *in D the sequence*

$$(22) \qquad \mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{M}_k^{-1}\mathbf{F}(\mathbf{x}_k), \qquad k = 0, 1, \ldots,$$

remains in D, $\mathbf{x}_k \neq \mathbf{x}_*$ *for* $k \geq 0$, *and converges to* $\mathbf{x}_*$. *Then* $\{\mathbf{x}_k\}$ *converges superlinearly in* $\mathbf{x}_*$ *if and only if*

$$(23) \qquad \lim_{k\to+\infty} \frac{\|[\mathbf{M}_k - \mathbf{J}_*](\mathbf{x}_{k+1} - \mathbf{x}_k)\|}{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|} = 0.$$

As noted in [8], if we define $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{s}_k^N = -\mathbf{J}_k^{-1}\mathbf{F}(\mathbf{x}_k)$ then it is possible to rephrase (23) without explicitly referring to an iteration matrix $\mathbf{M}_k$.

THEOREM 5. *Let assumptions* (A_1)–(A_3) *hold. Let* $\{\mathbf{x}_k\}$ *be a sequence of points that remains in D*, $\mathbf{x}_{k+1} \neq \mathbf{x}_k$ *and* $\mathbf{x}_k \neq \mathbf{x}_*$, *for* $k \geq 0$, *and converges to* $\mathbf{x}_*$. *Then* $\{\mathbf{x}_k\}$ *converges superlinearly to* $\mathbf{x}_*$ *if and only if*

$$(24) \qquad \lim_{k\to+\infty} \frac{\|\mathbf{s}_k - \mathbf{s}_k^N\|}{\|\mathbf{s}_k\|} = 0.$$

PROOF. It is easy to see that the assumptions of Theorem 4 imply the assumptions of Theorem 5. To see that the converse is true, assume that the assumptions of Theorem 5 hold. But we can define a nonsingular matrix mapping the vector $-\mathbf{F}(\mathbf{x}_k)$ to the vector $\mathbf{x}_{k+1} - \mathbf{x}_k$: Define $\mathbf{W}_k$ to be the matrix $\alpha \times \mathbf{Q}$ where $\mathbf{Q}$ is the orthogonal rotator that brings the vector $-\mathbf{F}(\mathbf{x}_k)$ onto the ray $(\mathbf{x}_{k+1} - \mathbf{x}_k)/\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ and define $\alpha = \|\mathbf{x}_{k+1} - \mathbf{x}_k\|/\|\mathbf{F}(\mathbf{x}_k)\|$. Obviously, $\mathbf{W}_k$ is well defined and nonsingular; let $\mathbf{M}_k = \mathbf{W}_k^{-1}$. The equivalence of the assumptions is established.

To complete the proof we must merely establish the equivalence of (23) and (24). But

$$\mathbf{s}_k - \mathbf{s}_k^N = \mathbf{J}_k^{-1}[\mathbf{J}_k - \mathbf{M}_k]\mathbf{s}_k$$

and the result follows. ∎

The application of (23) or (24) to constrained optimization is subtle: $\mathbf{F}$, as defined in (10), is a theoretical construction and is not, in general, computable at $\mathbf{x}_k$. A possible alternative is to derive a characterization that does not depend explicitly on $\mathbf{F}$. We do this next.

First we show that $\mathbf{J}_*$ can be replaced in (23).

Lemma 6. *Let the assumptions of Theorem 4 hold. Further, let $\{\mathbf{B}_k\}$ be any sequence of matrices satisfying*

(25) $$\lim_{k\to+\infty} \|\mathbf{B}_k - \mathbf{J}_*\| = 0.$$

Then $\{\mathbf{x}_k\}$ converges superlinearly to $\mathbf{x}_*$ if and only if

(26) $$\lim_{k\to+\infty} \frac{\|[\mathbf{M}_k - \mathbf{B}_k](\mathbf{x}_{k+1} - \mathbf{x}_k)\|}{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|} = 0.$$

Proof. We need only show that condition (26) is equivalent to (23). However,

(27) $$\frac{[\mathbf{M}_k - \mathbf{B}_k](\mathbf{x}_{k+1} - \mathbf{x}_k)}{(\mathbf{x}_{k+1} - \mathbf{x}_k)} = \frac{[\mathbf{M}_k - \mathbf{J}_*](\mathbf{x}_{k+1} - \mathbf{x}_k)}{(\mathbf{x}_{k+1} - \mathbf{x}_k)} + \frac{[\mathbf{J}_* - \mathbf{B}_k](\mathbf{x}_{k+1} - \mathbf{x}_k)}{(\mathbf{x}_{k+1} - \mathbf{x}_k)}$$

from which the result follows directly. ∎

Theorem 7. *Let the assumptions of Theorem 5 hold. Let $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and let $\mathbf{s}_k^{\mathrm{SQP}}$ be the* SQP *step (i.e.,* (3) *or* (6) *or* (14)*). Then $\{\mathbf{x}_k\}$ converges superlinearly to $\mathbf{x}_*$ if and only if*

(28) $$\lim_{k\to+\infty} \frac{\|\mathbf{s}_k - \mathbf{s}_k^{\mathrm{SQP}}\|}{\|\mathbf{s}_k\|} = 0.$$

Proof. The SQP step is the solution to a system $\mathbf{B}_k\mathbf{s}_k^{\mathrm{SQP}} = -\mathbf{F}(\mathbf{x}_k)$, where $\mathbf{B}_k$ and $\mathbf{F}$ are defined in (15) and (10), respectively; hence, by

Lemma 2, $\|\mathbf{B}_k - \mathbf{J}_*\| = O(\|\mathbf{x}_k - \mathbf{x}_*\|)$. But our assumptions imply the existence of a nonsingular matrix $\mathbf{M}_k$ such that $\mathbf{M}_k\mathbf{s}_k = -\mathbf{F}(\mathbf{x}_k)$ and therefore,

$$\mathbf{s}_k - \mathbf{s}_k^{\mathrm{SQP}} = \mathbf{B}_k^{-1}[\mathbf{B}_k - \mathbf{M}_k]\mathbf{s}_k \tag{29}$$

and the result follows using Lemma 6. ∎

Hence, a superlinear rate is achieved if and only if the steps approach the SQP steps, asymptotically, both in size and direction.

For completeness, we remark that the superlinearity characterization given in [6] can be expressed in the context of problem (1) using $\mathbf{F}$ given by (10).

THEOREM 8. *Let the assumptions of Theorem* 5 *hold. Then* $\mathbf{x}_k$ *converges to* $\mathbf{x}_*$ *superlinearly if and only if*

$$\lim_{k\to\infty} \frac{\left\| \begin{pmatrix} \mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla^2 L_k \\ \mathbf{A}_k^{\mathrm{T}} \end{pmatrix} \mathbf{s}_k + \begin{pmatrix} \mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla f_k \\ \mathbf{c}_k \end{pmatrix} \right\|}{\left\| \begin{pmatrix} \mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla f_k \\ \mathbf{c}_k \end{pmatrix} \right\|} = 0. \tag{30}$$

The proof is straightforward; we omit it. This type of characterization has proven useful in the unconstrained setting in the context of an iterative technique for solving the current linearized approximation inexactly; (30) is not immediately useful in the constrained setting because $\mathbf{Z}$ is not computable at $\mathbf{x}_k$, in general. However, if we define $\mathbf{P}_k^Z$ to be the orthogonal projector onto null($\mathbf{A}_k^{\mathrm{T}}$), then (30) is equivalent to

$$\lim_{k\to\infty} \frac{\left\| \begin{pmatrix} \mathbf{P}_k^Z(\nabla^2 L_k\mathbf{s}_k + \nabla f_k) \\ \mathbf{A}_k^{\mathrm{T}}\mathbf{s}_k + \mathbf{c}_k \end{pmatrix} \right\|}{\left\| \begin{pmatrix} \mathbf{P}_k^Z\nabla f_k \\ \mathbf{c}_k \end{pmatrix} \right\|} = 0 \tag{31}$$

which is computable using local information only. Note that (31) is independent of the choice of basis $\mathbf{Z}_k$.

3.1. *Quasi-Newton* SQP *methods.* A characterization of superlinear convergence for constrained optimization, in the special case when a quasi-Newton SQP algorithm is used, was first given in [1]. Theorem 4 is more general because it does not presuppose an algorithm class (also, linear convergence is assumed in [1], whereas convergence only is assumed above). However, the application of Theorem 4 is not obvious since $\mathbf{F}$—given by (7), (10)—is a theoretical device and is not, in general, computable at $\mathbf{x}_k$. Nevertheless, the restricted characterization in

[1] is a direct consequence of the following result which, in turn, is an easy consequence of Theorem 4.

LEMMA 9. *Let the assumptions of Theorem* 4 *hold. Further, let* $\{\mathbf{C}_k\}$ *be a sequence of matrices such that* $\{\|\mathbf{C}_k\|\}$ *is bounded above and*

$$\lim_{k\to\infty} \frac{\|(\mathbf{I}-\mathbf{C}_k)(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\|}{\|\mathbf{s}_k\|}=0. \tag{32}$$

Then $\{\mathbf{x}_k\}$ *converges to* $\mathbf{x}_*$ *superlinearly if and only if*

$$\lim_{k\to\infty} \frac{\|\mathbf{C}_k(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\|}{\|\mathbf{s}_k\|}=0. \tag{33}$$

PROOF. First suppose $\{\mathbf{x}_k\}$ converges superlinearly; therefore, by Theorem 4, (23) holds. But,

$$\|\mathbf{C}_k(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\| \le \|\mathbf{C}_k\|\cdot\|(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\| = O(\|(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\|)$$

and therefore (33) holds. On the other hand, suppose (33) is true. But,

$$\|(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\| \le \|(\mathbf{I}-\mathbf{C}_k)(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\| + \|\mathbf{C}_k(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\|,$$

which implies (23) and hence superlinear convergence. ■

It is worthwhile noting at this point that our application will involve the use of a specific choice for $\mathbf{C}_k$, constant for all k: $\mathbf{C}_k$ will be the orthogonal projector onto $\langle \mathbf{e}_1,\ldots,\mathbf{e}_{n-t}\rangle$, denoted by $\mathbf{P}$.

The following result was originally established by Boggs, Tolle, and Wang [1]. Since then alternative proofs have been provided in [9], [12], [13].

COROLLARY 10. *Let assumptions* (A_1)–(A_3) *hold. Let* $\mathbf{H}_k$ *be a symmetric matrix with the restriction of* $\mathbf{H}_k$ *onto* $\text{null}(\mathbf{A}_*^T)$ *positive definite* (*by restriction we mean the two-sided projection*). *Let* $\{\mathbf{x}_k\}$ *be defined by* $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{s}_k$, *where* $\mathbf{s}_k$ *solves*

$$\text{minimize}\left\{\mathbf{s}^T\nabla f_k + \tfrac{1}{2}\mathbf{s}^T\mathbf{H}_k\mathbf{s} \colon \mathbf{A}_k^T\mathbf{s} = -\mathbf{c}_k\right\}.$$

Assume $\{\mathbf{x}_k\} \to \mathbf{x}_*$; *let* $\mathbf{P}_k^Z$ *be the orthogonal projector onto* $\text{null}(\mathbf{A}_k^T)$. *Then* $\{\mathbf{x}_k\}$ *converges superlinearly to* $\mathbf{x}_*$ *if and only if*

$$\lim_{k\to+\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k-\nabla^2 L_*)\mathbf{s}_k\|}{\|\mathbf{s}_k\|}=0. \tag{35}$$

PROOF. For sufficiently small D the restriction of $\mathbf{H}_k$ onto $\text{null}(\mathbf{A}_k^T)$ is positive definite and therefore $\mathbf{s}_k$ is the solution to the system

$$\mathbf{M}_k\mathbf{s}_k = \begin{pmatrix} -\mathbf{Z}(\mathbf{x}_k)^T\nabla f_k \\ -\mathbf{c}_k \end{pmatrix} = -\mathbf{F}(\mathbf{x}_k), \tag{36}$$

where

$$\mathbf{M}_k = \begin{pmatrix} \mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{H}_k \\ \mathbf{A}_k^{\mathrm{T}} \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{M}_k^1 \\ \mathbf{M}_k^2 \end{pmatrix}, \tag{37}$$

$\mathbf{Z}(\mathbf{x}_k)$ is defined by (7), and $\mathbf{F}$ is given by (10). By (12),

$$\mathbf{J}_* = \begin{pmatrix} \mathbf{Z}_*^{\mathrm{T}}\nabla^2 L_* \\ \mathbf{A}_*^{\mathrm{T}} \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{J}_*^1 \\ \mathbf{J}_*^2 \end{pmatrix}. \tag{38}$$

Let $\mathbf{P}$ be the orthogonal projector onto $\langle \mathbf{e}_1, \dots, \mathbf{e}_{n-t} \rangle$; hence,

$$(\mathbf{I}-\mathbf{P})(\mathbf{M}_k - \mathbf{J}_*)\mathbf{s}_k = \begin{pmatrix} \mathbf{0} \\ (\mathbf{M}_k^2 - \mathbf{J}_*^2)\mathbf{s}_k \end{pmatrix}$$

and

$$\lim_{k\to\infty} \frac{\|(\mathbf{M}_k^2 - \mathbf{J}_*^2)\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = \lim_{k\to\infty} \frac{\|(\mathbf{A}_k^{\mathrm{T}} - \mathbf{A}_*^{\mathrm{T}})\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0 \tag{39}$$

and therefore assumption (32) of Lemma 9 is satisfied (with $\mathbf{C}_k = \mathbf{P}$ for all k). To complete the proof we must merely show the equivalence of (35) and (33) in this case where we restrict the algorithm to class (34). But,

$$\mathbf{P}(\mathbf{M}_k - \mathbf{J}_*)\mathbf{s}_k = \begin{pmatrix} (\mathbf{M}_k^1 - \mathbf{J}_*^1)\mathbf{s}_k \\ \mathbf{0} \end{pmatrix} \tag{40}$$

and

$$\begin{aligned} \|(\mathbf{M}_k^1 - \mathbf{J}_*^1)\mathbf{s}_k\| &= \|(\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{H}_k - \mathbf{Z}_*^{\mathrm{T}}\nabla^2 L_*)\mathbf{s}_k\| \\ &\le \|\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\| + \|(\mathbf{Z}(\mathbf{x}_k) - \mathbf{Z}_*)^{\mathrm{T}}\nabla^2 L_* \mathbf{s}_k\|. \end{aligned}$$

But,

$$\begin{aligned} &\|\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\| \\ &\quad = \|[\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)][\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)]^{-1}\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\| \\ &\quad \le \|\mathbf{Z}(\mathbf{x}_k)\| \cdot \|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\|. \end{aligned}$$

Therefore (35) implies (33). To see that (33) implies (35) use (40) and note that

$$\begin{aligned} \|(\mathbf{M}_k^1 - \mathbf{J}_*^1)\mathbf{s}_k\| &= \|(\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{H}_k - \mathbf{Z}_*^{\mathrm{T}}\nabla^2 L_*)\mathbf{s}_k\| \\ &= \|\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k + (\mathbf{Z}(\mathbf{x}_k) - \mathbf{Z}_*)^{\mathrm{T}}\nabla^2 L_* \mathbf{s}_k\| \\ &\ge \|\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\| - \|(\mathbf{Z}(\mathbf{x}_k) - \mathbf{Z}_*)^{\mathrm{T}}\nabla^2 L_* \mathbf{s}_k\|. \end{aligned}$$

But

$$\begin{aligned} \|\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\| &= \frac{\|\mathbf{Z}(\mathbf{x}_k)[\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)]^{-1}\|}{\|\mathbf{Z}(\mathbf{x}_k)[\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)]^{-1}\|} \\ &\qquad \cdot \|\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\| \\ &\ge \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\|}{\|\mathbf{Z}(\mathbf{x}_k)[\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)]^{-1}\|}. \end{aligned}$$

Hence, (33) implies (35) and the proof is complete. ∎

Next we make additional restrictions on the approximating matrix $\mathbf{M}_k$.

LEMMA 11. *Let the assumptions of Theorem 4 hold; assume that* $\{\mathbf{C}_k^L\}$ *and* $\{\mathbf{C}_k^R\}$ *are sequences of matrices such that* $\{\|\mathbf{C}_k^L\|\}$ *is bounded from above. Further, assume*

$$\lim_{k\to+\infty} \frac{\|(\mathbf{I}-\mathbf{C}_k^L)(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0 \tag{41}$$

and

$$\lim_{k\to+\infty} \frac{\|\mathbf{C}_k^L(\mathbf{M}_k-\mathbf{J}_*)(\mathbf{I}-\mathbf{C}_k^R)\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0. \tag{42}$$

Then $\{\mathbf{x}_k\}$ *converges to* $\mathbf{x}_*$ *superlinearly if and only if*

$$\lim_{k\to+\infty} \frac{\|\mathbf{C}_k^L(\mathbf{M}_k-J_*)\mathbf{C}_k^R\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0. \tag{43}$$

PROOF. First suppose $\{\mathbf{x}_k\}$ converges superlinearly to $\mathbf{x}_*$. But

$$\begin{aligned}\|\mathbf{C}_k^L(\mathbf{M}_k-\mathbf{J}_*)\mathbf{C}_k^R\mathbf{s}_k\| &\le \|\mathbf{C}_k^L(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\| + \|\mathbf{C}_k^L(\mathbf{M}_k-\mathbf{J}_*)(\mathbf{I}-\mathbf{C}_k^R)\mathbf{s}_k\| \\ &\le \|\mathbf{C}_k^L\|\cdot\|(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k\| \\ &\qquad + \|\mathbf{C}_k^L(\mathbf{M}_k-\mathbf{J}_*)(\mathbf{I}-\mathbf{C}_k^R)\mathbf{s}_k\|\end{aligned}$$

and therefore by Theorem 4 and assumption (42), (43) follows. Next assume (43). But

$$\begin{aligned}(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k &= (\mathbf{I}-\mathbf{C}_k^L)(\mathbf{M}_k-\mathbf{J}_*)\mathbf{s}_k \\ &\quad + \mathbf{C}_k^L(\mathbf{M}_k-\mathbf{J}_*)\mathbf{C}_k^R\mathbf{s}_k + \mathbf{C}_k^L(\mathbf{M}_k-\mathbf{J}_*)(\mathbf{I}-\mathbf{C}_k^R)\mathbf{s}_k\end{aligned}$$

and therefore by Theorem 4 superlinear convergence follows. ∎

We can apply Lemma 11 to a quasi-Newton SQP-algorithm by making specific choices for matrices $\mathbf{C}_k^L$ and $\mathbf{C}_k^R$.

THEOREM 12. *Let the assumptions of Corollary 10 hold; further, assume*

$$\lim_{k\to\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k-\nabla^2 L_*)(\mathbf{I}-\mathbf{P}_k^Z)\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0. \tag{44}$$

Then $\{\mathbf{x}_k\}$ *converges superlinearly to* $\mathbf{x}_*$ *if and only if*

$$\lim_{k\to\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k-\nabla^2 L_*)\mathbf{P}_k^Z\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0. \tag{45}$$

Proof. From Corollary 10 it follows that superlinear convergence is achieved if and only if

$$(46)\qquad \lim_{k\to+\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0.$$

But,

$$(47)\quad \mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k = \mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)(\mathbf{I} - \mathbf{P}_k^Z)\mathbf{s}_k + \mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{P}_k^Z\mathbf{s}_k.$$

The result now follows trivially. ∎

Theorem 12 has interesting algorithmic implications. Specifically, it turns out that it is possible to satisfy assumption (44) while only maintaining an explicit approximation to $\mathbf{P}_k^Z\nabla^2 L_k \mathbf{P}_k^Z$. To see this consider the following (two-step) algorithm, where $\mathbf{H}_k$ is a symmetric matrix, of order n, and positive definite on null$(\mathbf{A}_k^T)$.

Solve

$$(48)\qquad \begin{pmatrix} \mathbf{Z}_k^T\mathbf{H}_k\mathbf{Z}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_k^T \end{pmatrix} \begin{pmatrix} \mathbf{h}_k \\ \mathbf{v}_k \end{pmatrix} = \begin{pmatrix} -\mathbf{Z}_k^T(\nabla L_k(\mathbf{x}_k + \mathbf{Y}_k\mathbf{v}_k)) \\ -\mathbf{c}_k \end{pmatrix}$$

where $\mathbf{Z}_k$ is any basis for null$(\mathbf{A}_k^T)$, and then set

$$(49)\qquad \mathbf{s}_k \leftarrow \mathbf{Z}_k\mathbf{h}_k + \mathbf{Y}_k\mathbf{v}_k, \quad \mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{s}_k.$$

Corollary 13. *Let assumptions* (A_1)–(A_3) *hold. Let* $\mathbf{H}_k$ *be a symmetric matrix with the restriction of* $\mathbf{H}_k$ *onto* null$(\mathbf{A}_*^T)$ *positive definite (by restriction we mean the two-sided projection). Let* $\{\mathbf{x}_k\}$ *be defined by* (48)–(49). *Assume that* $\{\lambda_k\} \to \lambda_*$; $\{\mathbf{x}_k\} \to \mathbf{x}_*$. *Let* $\mathbf{P}_k^Z$ *be the orthogonal projector onto* null$(\mathbf{A}_k^T)$. *Then* $\{\mathbf{x}_k\}$ *converges superlinearly to* $\mathbf{x}_*$ *if and only if* (45) *holds.*

Proof. By Taylor's theorem, for any vector $\mathbf{u}_k$,

$$\nabla L_k(\mathbf{x}_k + \mathbf{u}_k) = \nabla L_k(\mathbf{x}_k) + \nabla^2 L_k(\mathbf{x}_k)\mathbf{u}_k + \mathbf{w}_k$$

where $\|\mathbf{w}_k\| = o(\|\mathbf{u}_k\|)$. Equivalently,

$$(50)\qquad (\nabla^2 L_k(\mathbf{x}_k) + \mathbf{E}_k)\mathbf{u}_k = \nabla L_k(\mathbf{x}_k + \mathbf{u}_k) - \nabla L_k(\mathbf{x}_k)$$

where $\mathbf{E}_k = \mathbf{w}_k(\mathbf{u}_k^T\mathbf{u}_k)^+\mathbf{u}_k^T$ and $(\cdot)^+$ denotes the pseudo-inverse (i.e., $\alpha^+ = \alpha^{-1}$ if $\alpha \neq 0$; otherwise, $\alpha^+ = 0$). But

$$\|\mathbf{E}_k\| \leq \|\mathbf{w}_k\| \cdot \|(\mathbf{u}_k^T\mathbf{u}_k)^+\mathbf{u}_k^T\|$$

and therefore, if $\mathbf{u}_k = 0$ then $\mathbf{E}_k = \mathbf{0}$; otherwise,

$$\|\mathbf{E}_k\| \leq \frac{o(\|\mathbf{u}_k\|)}{\|\mathbf{u}_k\|}$$

and so

$$\text{if } \{\|\mathbf{u}_k\|\} \to 0 \quad \text{then } \{\|\mathbf{E}_k\|\} \to 0. \tag{51}$$

If $\mathbf{u}_k = \mathbf{Y}_k\mathbf{v}_k$ then by (50) we can rewrite (48)–(49) as

$$\begin{pmatrix} \mathbf{Z}_k^T\mathbf{H}_k\mathbf{Z}_k & \mathbf{Z}_k^T(\nabla^2 L_k(\mathbf{x}_k) + \mathbf{E}_k)\mathbf{Y}_k \\ \mathbf{0} & \mathbf{R}_k^T \end{pmatrix} \begin{pmatrix} \mathbf{h}_k \\ \mathbf{v}_k \end{pmatrix} = \begin{pmatrix} -\mathbf{Z}_k^T\nabla f_k \\ -\mathbf{c}_k \end{pmatrix} \tag{52}$$

and then set

$$\mathbf{s}_k \leftarrow \mathbf{Z}_k\mathbf{h}_k + \mathbf{Y}_k\mathbf{v}_k, \quad \mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{s}_k. \tag{53}$$

But (52)–(53) can be expressed in the form (34) using matrix $\widehat{\mathbf{H}}_k$, where

$$\widehat{\mathbf{H}}_k = \mathbf{P}_k^Z\mathbf{H}_k\mathbf{P}_k^Z + \mathbf{P}_k^Z(\nabla^2 L_k(\mathbf{x}_k) + \mathbf{E}_k)(\mathbf{I} - \mathbf{P}_k^Z). \tag{54}$$

However,

$$\mathbf{P}_k^Z(\widehat{\mathbf{H}}_k - \nabla^2 L_*)(\mathbf{I} - \mathbf{P}_k^Z) = \mathbf{P}_k^Z(\nabla^2 L_k(\mathbf{x}_k) + \mathbf{E}_k - \nabla^2 L_*)(\mathbf{I} - \mathbf{P}_k^Z)$$

and therefore, using (51) and $\{\lambda_k\} \to \lambda_*$, assumption (44) of Theorem 12 is satisfied and the result follows. ■

This last method is an *approximate* quasi-Newton SQP method: $\mathbf{s}_k$ is not computed by solving a problem of the form (34). However, we include it in this subsection because theoretically it is easily expressed in this form (as we have seen). This is not the case for the class of methods described next.

3.2. *Approximate quasi-Newton* SQP *methods.* Unfortunately, the characterizations of superlinear convergence for quasi-Newton SQP methods are not as useful as one would hope: existing quasi-Newton methods do not always fit precisely into the SQP mode. However, it is possible to view such methods as *approximate* quasi-Newton SQP procedures; the (unrestricted) superlinear characterizations, discussed at the beginninig of this section, can be applied to establish a superlinear characterization for a broad class of *approximate* quasi-Newton SQP methods. We do this next.

The following algorithm uses, at each step, matrices $\widehat{\mathbf{Z}}_k$ and $\widehat{\mathbf{A}}_k$; note that, in general, $\langle\widehat{\mathbf{Z}}_k\rangle \neq \langle\mathbf{Z}(\mathbf{x}_k)\rangle$, $\langle\widehat{\mathbf{A}}_k\rangle \neq \langle\mathbf{A}(\mathbf{x}_k)\rangle$, and $\widehat{\mathbf{A}}_k^T\widehat{\mathbf{Z}}_k \neq 0$.[2] Let $\widehat{\mathbf{A}}_k = \widehat{\mathbf{Y}}_k\widehat{\mathbf{R}}_k$.

[2]If $\mathbf{M}$ is a matrix then $\langle\mathbf{M}\rangle$ refers to the space spanned by the columns of $\mathbf{M}$.

Solve

$$\begin{pmatrix} \hat{\mathbf{Z}}_k^{\mathrm{T}}\mathbf{H}_k\hat{\mathbf{Z}}_k & \hat{\mathbf{Z}}_k^{\mathrm{T}}\mathbf{H}_k\hat{\mathbf{Y}}_k \\ \mathbf{0} & \hat{\mathbf{R}}_k^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \mathbf{h}_k \\ \mathbf{v}_k \end{pmatrix} = \begin{pmatrix} -\hat{\mathbf{Z}}_k^{\mathrm{T}}\nabla L_k(\mathbf{x}_k) \\ -\mathbf{c}_k \end{pmatrix}, \tag{55}$$

$$\mathbf{s}_k \leftarrow \hat{\mathbf{Z}}_k\mathbf{h}_k + \hat{\mathbf{Y}}_k\mathbf{v}_k, \tag{56}$$

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{s}_k. \tag{57}$$

THEOREM 14. *Let assumptions* (A_1)–(A_3) *hold. Let* $\{\mathbf{s}_k\}$ *and* $\{\mathbf{x}_k\}$ *be generated as above. Assume* $\{\mathbf{x}_k\} \to \mathbf{x}_*$. *Further, assume*

(i) *there exist matrices* $\mathbf{S}_k$ *and* $\mathbf{T}_k$ *such that*

$$\hat{\mathbf{Z}}_k = \mathbf{Z}(\mathbf{x}_k)\mathbf{T}_k + \mathbf{S}_k$$

where the singular values of $\mathbf{T}_k$ *are bounded below away from zero and bounded above, and*

$$\|\mathbf{S}_k\| = O(\|\mathbf{s}_k\|);$$

(ii) $\lim_{k\to\infty} \hat{\mathbf{A}}_k = \mathbf{A}_*$;

(iii) $\{\mathbf{H}_k\}$ *is a sequence of matrices such that the restriction of* $\mathbf{H}_k$ *onto* $\mathrm{null}(\mathbf{A}_*^{\mathrm{T}})$ *is positive definite;* $\{\|\mathbf{H}_k\|\}$ *is bounded above;*

(iv) $\lim_{k\to\infty}\{\lambda_k\} = \lambda_*$.

Then, $\{\mathbf{x}_k\}$ *converges superlinearly to* $\mathbf{x}_*$ *if and only if*

$$\lim_{k\to+\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0. \tag{58}$$

PROOF. Our proof technique is to establish that $\mathbf{s}_k$ solves a system

$$\begin{pmatrix} \mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{H}_k + \mathbf{E}_k^1 \\ \hat{\mathbf{A}}_k^{\mathrm{T}} + \mathbf{E}_k^2 \end{pmatrix} \mathbf{s}_k = \begin{pmatrix} -\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla f_k \\ -\mathbf{c}_k \end{pmatrix} \tag{59}$$

such that $\lim_{k\to\infty}\|\mathbf{E}_k^1\| = \lim_{k\to\infty}\|\mathbf{E}_k^2\| = 0$, and $\mathbf{Z}(\mathbf{x}_k)$ is defined by (7). The result then follows directly from Lemma 9 (with $\mathbf{C}_k = \mathbf{P}$, the orthogonal projector onto $\langle \mathbf{e}_1, \ldots, \mathbf{e}_{n-t}\rangle$).

But (55)–(57) can be written

$$\begin{pmatrix} \hat{\mathbf{Z}}_k^{\mathrm{T}}\mathbf{H}_k \\ \hat{\mathbf{A}}_k^{\mathrm{T}} + \mathbf{E}_k^2 \end{pmatrix} \mathbf{s}_k = \begin{pmatrix} -\hat{\mathbf{Z}}_k^{\mathrm{T}}\nabla L_k \\ -\mathbf{c}_k \end{pmatrix} \tag{60}$$

where $\mathbf{E}_k^2 = -\mathbf{w}_k^2\mathbf{s}_k^{\mathrm{T}}/(\mathbf{s}_k^{\mathrm{T}}\mathbf{s}_k)$ and $\mathbf{w}_k^2 = \hat{\mathbf{A}}_k^{\mathrm{T}}\hat{\mathbf{Z}}_k\mathbf{h}_k$. Note that

$$\|\mathbf{E}_k^2\| \le \frac{\|\mathbf{w}_k^2\|}{\|\mathbf{s}_k\|} = O(\|\hat{\mathbf{A}}_k^{\mathrm{T}}\hat{\mathbf{Z}}_k\|)$$

and therefore, by convergence and assumptions (i) and (ii), $\lim_{k\to\infty}\|\mathbf{E}_k^2\| = 0$.

Next, using (i), we write

$$\hat{\mathbf{Z}}_k = \mathbf{Z}(\mathbf{x}_k)\mathbf{T}_k + \mathbf{S}_k$$

where $\{\mathbf{T}_k\}$ is uniformly nonsingular and $\|\mathbf{S}_k\| = O(\|\mathbf{s}_k\|)$. Hence,

$$\hat{\mathbf{Z}}_k^{\mathrm{T}}\nabla L_k = \mathbf{T}_k^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla L_k + \mathbf{w}_k^1$$

where $\mathbf{w}_k^1 \overset{\text{def}}{=} \mathbf{S}_k^{\mathrm{T}}\nabla L_k$. If we define $\tilde{\mathbf{E}}_k^1 = \mathbf{w}_k^1\mathbf{s}_k^{\mathrm{T}}/(\mathbf{s}_k^{\mathrm{T}}\mathbf{s}_k)$ then (60) can be written

$$\begin{pmatrix} \hat{\mathbf{Z}}_k^{\mathrm{T}}\mathbf{H}_k + \tilde{\mathbf{E}}_k^1 \\ \hat{\mathbf{A}}_k^{\mathrm{T}} + \mathbf{E}_k^2 \end{pmatrix} \mathbf{s}_k = \begin{pmatrix} -\mathbf{T}_k^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\nabla L_k \\ -\mathbf{c}_k \end{pmatrix}. \tag{61}$$

Note that $\|\tilde{\mathbf{E}}_k^1\| \le \|\mathbf{w}_k^1\|/\|\mathbf{s}_k\|$ and

$$\|\mathbf{w}_k^1\| \le \|\mathbf{S}_k\| \cdot \|\nabla L_k\|. \tag{62}$$

But, by assumption (i), $\|\mathbf{S}_k\| = O(\|\mathbf{s}_k\|)$ and $\nabla L_k \to \nabla L_* = 0$, using (iv); therefore, $\lim_{k\to\infty} \|\tilde{\mathbf{E}}_k^1\| = 0$.

Finally,

$$\begin{aligned} \hat{\mathbf{Z}}_k^{\mathrm{T}}\mathbf{H}_k &= \mathbf{T}_k^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{H}_k + \mathbf{S}_k^{\mathrm{T}}\mathbf{H}_k \\ &= \mathbf{T}_k^{\mathrm{T}}\mathbf{Z}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{H}_k + \mathbf{E}_k^3 \end{aligned}$$

where $\mathbf{E}_k^3 \overset{\text{def}}{=} \mathbf{S}_k^{\mathrm{T}}\mathbf{H}_k$ and therefore, using assumptions (iii) and (i), $\lim_{k\to\infty} \|\mathbf{E}_k^3\| = 0$. System (59) is now obtained with $\mathbf{E}_k^1 = \mathbf{T}_k^{-\mathrm{T}}(\tilde{\mathbf{E}}_k^1 + \mathbf{E}_k^3)$. ■

Note. The assumption that $\|\mathbf{S}_k\| = O(\|\mathbf{s}_k\|)$ is practical since it is satisfied by most (if not all) known superlinearly convergent updating schemes. However, it is possible to replace this assumption with conditions on $\{\mathbf{x}_k\}$ and $\{\lambda_k\}$. Specifically, we can replace assumption (i) with (i′):

(i′) There exist matrices $\mathbf{S}_k$ and $\mathbf{T}_k$ such that

$$\hat{\mathbf{Z}}_k = \mathbf{Z}(\mathbf{x}_k)\mathbf{T}_k + \mathbf{S}_k$$

where the singular values of $\mathbf{T}_k$ are bounded below away from zero and bounded above, and

$$\|\mathbf{S}_k\| \to 0.$$

Moreover, we assume that $\{\mathbf{x}_k\}$ converges to $\mathbf{x}_*$ *linearly* and $\|\lambda_k - \lambda_*\| = O(\|\mathbf{x}_k - \mathbf{x}_*\|)$.

Then (62) still holds but, using Taylor's theorem,

(63)
$$\|\nabla L_k\| = O(\|\mathbf{x}_k - \mathbf{x}_*\|) + O(\|\lambda_k - \lambda_*\|)$$
(64) $$= O(\|\mathbf{x}_k - \mathbf{x}_*\|) \quad (by\ (\mathrm{i}'))$$
(65) $$= O(\|\mathbf{s}_k\|) \quad (by\ the\ linear\ convergence\ assumption\ in\ (\mathrm{i}')).$$

Consequently, $\lim_{k\to\infty} \|\widehat{\mathbf{E}}_k^1\| = 0$ and the rest of the argument follows unchanged.

If, in addition, we assume

(66) $$\lim_{k\to\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2\mathbf{L}_*)(\mathbf{I} - \mathbf{P}_k^Z)\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0,$$

then the remaining requirement for a superlinear characterization is on $\mathbf{P}_k^Z\mathbf{H}_k\mathbf{P}_k^Z$ alone.

THEOREM 15. *Let the assumptions of Theorem* 14 *hold. Further, assume* (66). *Then* $\{\mathbf{x}_k\}$ *converges to* $\mathbf{x}_*$ *superlinearly if and only if*

(67) $$\lim_{k\to\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{P}_k^Z\mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0.$$

PROOF. Assume superlinear convergence to $\mathbf{x}_*$. But,

$$\|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_k)\mathbf{P}_k^Z\mathbf{s}_k\| \le \|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_k)\mathbf{s}_k\| + \|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_k)(\mathbf{I} - \mathbf{P}_k^Z)\mathbf{s}_k\|$$

and so by (58) and assumption (66), (67) follows. On the other hand, assume (67). But,

$$\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_k)\mathbf{s}_k = \mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_k)(\mathbf{I} - \mathbf{P}_k^Z)\mathbf{s}_k + \mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_k)\mathbf{P}_k^Z\mathbf{s}_k.$$

Hence, (66) and (67) yield (58) and by Theorem 14 superlinear convergence is established. ■

It is possible to satisfy assumption (66) using an extra gradient evaluation. Specifically, let us replace algorithm (55)–(57) with

Solve

(68) $$\begin{pmatrix} \widehat{\mathbf{Z}}_k^{\mathrm{T}}\mathbf{H}_k\widehat{\mathbf{Z}}_k & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{R}}_k^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \mathbf{h}_k \\ \mathbf{v}_k \end{pmatrix} = \begin{pmatrix} -\widehat{\mathbf{Z}}_k^{\mathrm{T}}\nabla L_k(\mathbf{x}_k + \widehat{\mathbf{Y}}_k\mathbf{v}_k) \\ -\mathbf{c}_k \end{pmatrix},$$

(69) $$\mathbf{s}_k \leftarrow \widehat{\mathbf{Z}}_k\mathbf{h}_k + \widehat{\mathbf{Y}}_k\mathbf{v}_k,$$

(70) $$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{s}_k.$$

THEOREM 16. *Let the assumptions of Theorem* (14) *hold with the exception that algorithm* (55)–(57) *is replaced with* (68)–(70). *Then,* $\{\mathbf{x}_k\}$ *converges to* $\mathbf{x}_*$ *superlinearly if and only if*

$$\lim_{k\to\infty} \frac{\|\mathbf{P}_k^Z(\mathbf{H}_k - \nabla^2 L_*)\mathbf{P}_k^Z \mathbf{s}_k\|}{\|\mathbf{s}_k\|} = 0. \tag{71}$$

PROOF. Repeating the technique used in the proof of Corollary 13, we obtain a matrix $\mathbf{G}_k$ such that

$$\|\mathbf{G}_k\| \le \frac{o(\|\widehat{\mathbf{Y}}\mathbf{v}_k\|)}{\|\widehat{\mathbf{Y}}\mathbf{v}_k\|}$$

and (68)–(70) is equivalent to

Solve

$$\begin{pmatrix} \widehat{\mathbf{Z}}_k^{\mathrm{T}}\mathbf{H}_k\widehat{\mathbf{Z}}_k & \widehat{\mathbf{Z}}_k^{\mathrm{T}}(\nabla^2 L_k(\mathbf{x}_k) + \mathbf{G}_k)\widehat{\mathbf{Y}}_k \\ \mathbf{0} & \widehat{\mathbf{R}}_k^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \mathbf{h}_k \\ \mathbf{v}_k \end{pmatrix} = \begin{pmatrix} -\widehat{\mathbf{Z}}_k^{\mathrm{T}}\nabla L_k \\ -\mathbf{c}_k \end{pmatrix}. \tag{72}$$

Therefore, if we define

$$\widetilde{\mathbf{H}}_k = \mathbf{P}_k^Z\mathbf{H}_k\mathbf{P}_k^Z + \mathbf{P}_k^Z(\nabla^2 L_k(\mathbf{x}_k) + \mathbf{G}_k)(\mathbf{I} - \mathbf{P}_k^Z),$$

then (68)–(70) can be expressed in the form of algorithm (55)–(57 using $\widetilde{\mathbf{H}}_k$ to play the role of the matrix $\mathbf{H}_k$ in (55)–(57). The result no follows easily. ∎

4. An application. In this section we illustrate the usefulness of th viewpoint developed in the previous sections by providing a new pro of superlinear convergence of the constrained quasi-Newton method du to Coleman and Conn [4]. Coleman and Conn [4] established two-ste superlinear convergence; subsequently, Byrd [2] strengthened the resu to (one-step) superlinear convergence. Byrd's proof is quite differe from the proof we present here.

4.1. *The Coleman-Conn algorithm.* The algorithm recurs a positi definite matrix of order $n - t$, $\overline{\mathbf{H}}_k$. Let $\mathbf{x}_{k\_}$ be a previous point wi constraint matrix $\mathbf{A}_{k\_} = \mathbf{Y}_{k\_}\mathbf{R}_{k\_}$. The columns of the matrix $\overline{\mathbf{Z}}($ form an orthonormal basis for the null space of $\mathbf{A}(\mathbf{x})^{\mathrm{T}}$. The functi UPDATE($\mathbf{M}, \mathbf{s}, \mathbf{y}$) refers to either of the well-known positive defin secant updates, BFGS or DFP: the matrix $\mathbf{M}$ is a positive definite matr $\mathbf{s}$ is the current "step," $\mathbf{y}$ is the difference in "gradients," and $\mathbf{s}^{\mathrm{T}}\mathbf{y} > 0$

THE ALGORITHM.

(1) *Solve* $\mathbf{R}_{k_-}^T \mathbf{v}_k = -\mathbf{c}_k$.
(2) $\mathbf{x}_{k_+} \leftarrow \mathbf{x}_k + \mathbf{Y}_{k_-}\mathbf{v}_k$
(3) *Compute* $\mathbf{A}_{k_+} = \mathbf{Y}_{k_+}\mathbf{R}_{k_+}$; *compute* $\overline{\mathbf{Z}}_{k_+}$, *an orthonormal basis for* $\text{null}(\mathbf{A}_{k_+}^T)$.
(4) *Compute* $\nabla f_{k_+} \stackrel{\text{def}}{=} \nabla f(\mathbf{x}_{k_+})$.
(5) *Solve* $\mathbf{R}_{k_+}\lambda_{k_+} = \mathbf{Y}_{k_+}^T \nabla f_{k_+}$.
(6) *Solve* $\overline{\mathbf{H}}_k \mathbf{h}_k = -\overline{\mathbf{Z}}_{k_+}^T \nabla f_{k_+}$.
(7) $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_{k_+} + \overline{\mathbf{Z}}_{k_+}\mathbf{h}_k$
(8) $\overline{\mathbf{H}}_{k+1} \leftarrow \text{UPDATE}\{\overline{\mathbf{H}}_k, \mathbf{h}_k, \overline{\mathbf{Z}}_{k_+}^T(\nabla f_{k+1} - \mathbf{A}_{k+1}\lambda_{k_+}) - \overline{\mathbf{Z}}_{k_+}^T \nabla f_{k_+}\}$
(9) $k \leftarrow k+1,\ k_- \leftarrow k_+$

The key to the proof of superlinearity is that the correction to $\mathbf{x}_k$ can be expressed in the form (68)–(70). To see this note that $\mathbf{x}_{k+1}$ can be obtained via:

Solve

$$\begin{pmatrix} \overline{\mathbf{H}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{k_-} \end{pmatrix} \begin{pmatrix} \mathbf{h}_k \\ \mathbf{v}_k \end{pmatrix} = \begin{pmatrix} -\overline{\mathbf{Z}}_{k_+}^T \nabla L_k(\mathbf{x}_k + \mathbf{Y}_{k_-}\mathbf{v}_k) \\ -\mathbf{c}_k \end{pmatrix}, \tag{73}$$

$$\mathbf{s}_k \leftarrow \overline{\mathbf{Z}}_{k_+}\mathbf{h}_k + \mathbf{Y}_{k_-}\mathbf{v}_k, \tag{74}$$

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{s}_k. \tag{75}$$

Therefore, if we let $\mathbf{H}_k = \overline{\mathbf{Z}}_{k_+}\overline{\mathbf{H}}_k\overline{\mathbf{Z}}_{k_+}^T$ we see that the Coleman-Conn algorithm is in the form (68)–(70).

THEOREM 17. *Let assumptions* (A_1)–(A_3) *hold. Let* $\{\mathbf{x}_k\}$ *and* $\{\mathbf{H}_k\}$ *be generated as described above with* $\{\mathbf{x}_0, \overline{\mathbf{H}}_0\}$ *the starting pair,* $\mathbf{x}_{0_-} = \mathbf{x}_0$, *and* $\overline{\mathbf{H}}_0$ *is symmetric positive definite. Further, assume that* $\overline{\mathbf{Z}}(\mathbf{x})$ *is a Lipschitz continuous function on D; the Hessian matrices* $\nabla^2 f$ *and* $\nabla^2 c_i$, $i = 1, \ldots, t$, *are Lipschitz continuous on D.*

Then, there exist positive scalars ε *and* Δ *such that if* $\|\mathbf{x}_0 - \mathbf{x}_*\| \le \varepsilon$ *and* $\|\mathbf{H}_0 - \mathbf{M}_*\| \le \Delta$, *where* $\mathbf{M}_* = \overline{\mathbf{Z}}_*^T(\nabla^2 f_* + \sum \lambda_i^* \nabla^2 c_i^*)\overline{\mathbf{Z}}_*$ *and* $\overline{\mathbf{Z}}_*$ *is the limit point of* $\{\overline{\mathbf{Z}}_k\}$, *then* $\{\mathbf{x}_k\}$ *converges at a superlinear rate.*

PROOF. Since Coleman and Conn [4] have established convergence, we can prove superlinearity by applying Theorem 16 provided we establish that the assumptions of Theorem 14 hold. But assumptions (ii)–(iv) of Theorem 14 follow straightforwardly (with the boundedness of $\{\|\overline{\mathbf{H}}_k\|\}$ given by Coleman and Conn [4]). To establish that assumption

(i) of Theorem 14 holds, note that we can write

$$\mathbf{Z}(\mathbf{x}_{k_+}) = \mathbf{Z}(\mathbf{x}_k) + \left(\int_0^1 \dot{\mathbf{Z}}(\mathbf{x}_k + \tau(\mathbf{x}_{k_+} - \mathbf{x}_k))\partial\tau\right) \cdot (\mathbf{x}_{k_+} - \mathbf{x}_k). \tag{76}$$

But $\|\dot{\mathbf{Z}}(\mathbf{x})\|$ is bounded above for sufficiently small D, and therefore

$$\mathbf{Z}(\mathbf{x}_{k_+}) = \mathbf{Z}(\mathbf{x}_k) + \widehat{\mathbf{S}}_k \tag{77}$$

where $\widehat{\mathbf{S}}_k = O(\|\mathbf{s}_k\|)$. Since both $\overline{\mathbf{Z}}_{k_+}$ and $\mathbf{Z}(x_{k_+})$ are bases for null($\mathbf{A}_{k_+}^{\mathrm{T}}$) there exists a nonsingular matrix $\mathbf{T}_k$ such that

$$\overline{\mathbf{Z}}_{k_+} = \mathbf{Z}(\mathbf{x}_{k_+})\mathbf{T}_k. \tag{78}$$

In fact, continuity of $\overline{\mathbf{Z}}_k$ and $\mathbf{Z}(\mathbf{x}_k)$ and convergence yield that th columns of $\mathbf{T}_k$ are uniformly linearly independent and $\{\|\mathbf{T}_k\|\}$ is bound ed above, for sufficiently small D. Using (77) and (78), we have

$$\overline{\mathbf{Z}}_{k_+} = \mathbf{Z}(\mathbf{x}_k)\mathbf{T}_k + \mathbf{S}_k \tag{79}$$

where $\mathbf{S}_k = \widehat{\mathbf{S}}_k\mathbf{T}_k$ and $\|\mathbf{S}_k\| = O(\|\mathbf{s}_k\|)$. Therefore, assumption (i) o Theorem 14 is established.

Next an argument similar to that used to establish Theorem 3.6 an Corollary 3.11 of Coleman and Conn [4] can be used to give

$$\frac{\|(\overline{\mathbf{H}}_k - \mathbf{M}_*)\mathbf{h}_k\|}{\|\mathbf{s}_k\|} \to 0. \tag{80}$$

But (80) implies (71); the theorem is established.

4.2. *Concluding remarks.* (1) The two-step superlinear convergenc result of Coleman and Conn [4] was actually with respect to the sequenc $\{\mathbf{x}_{k_+}\}$ as defined above. It was Richard Byrd who first suggested—an later proved [2]—that the $\{\mathbf{x}_k\}$ sequence might be (one-step) superlir ear.

(2) The proof of (80) does depend on the use of a smooth $\overline{\mathbf{Z}}(\mathbf{x})$; as w have mentioned, this can be a thorny issue. Practical constructions hav been suggested in [5, 10]; however, Byrd and Schnabel [3] proposed modification to the Coleman–Conn algorithm that is *not* dependent o the choice of basis. The Byrd–Schnabel algorithm can be expressed i the same form as above with an additional "adjustment" to $\overline{\mathbf{H}}_k$ to refle the change in $\mathbf{Z}$. Specifically, precede step (6) with

$$\mathbf{H}_k \leftarrow \mathbf{T}_k^{\mathrm{T}}\mathbf{H}_k\mathbf{T}_k + \mathbf{Z}_{k_+}^{\mathrm{T}}\mathbf{Y}_{k_-}\overline{\mathbf{C}}_k\mathbf{Y}_{k_-}^{\mathrm{T}}\mathbf{Z}_{k_+} \tag{6$^-$}$$

where $\mathbf{T}_k = \mathbf{Z}_{k_-}^{\mathrm{T}}\mathbf{Z}_{k_+}$. This is actually a slight generalization of the Byrd Schnabel suggestion: they proposed what amounts to a specific choi for $\overline{\mathbf{C}}_k$, $\overline{\mathbf{C}}_k = \beta_k \mathbf{I}_t$, where β_k is a scale factor. Here, $\overline{\mathbf{C}}_k$ is symmetric but otherwise arbitrary. (The idea is that the matrix $\mathbf{Z}_k \mathbf{H}_k \mathbf{Z}_k^T + \mathbf{Y}_k \mathbf{C}_k \mathbf{Y}_k^T$ represents an approximation to the Hessian of the Lagrangian.)

Convergence properties for this algorithm are unknown; our development here suggests a possible way to proceed. Specifically, because the algorithm can be expressed in the form (68)–(70), a major step toward a proof is to show that (71) holds. It appears that a bounded deterioration result could be used, similar to the manner suggested by Coleman and Conn [4], provided the perturbation introduced by (6^-) is bounded by $O(\max\{\|\mathbf{x}_k - \mathbf{x}_*\|, \|\mathbf{x}_{k+1} - \mathbf{x}_*\|\})$.

References

1. P. T. Boggs, J. W. Tolle, and P. Wang, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM Journal on Control and Optimization **20** (1982), 161–171.

2. R. H. Byrd, *On the convergence of constrained optimization methods with accurate Hessian information on a subspace*, Tech. Rep. CU-CS-270-84, Department of Computer Science, University of Colorado (1984).

3. R. H. Byrd and R. B. Schnabel, *Continuity of the null space basis and constrained optimization*, Mathematical Programming **35** (1986), 32–41.

4. T. F. Coleman and A. R. Conn, *On the local convergence of a quasi-Newton method for the nonlinear programming problem*, SIAM J. Num. Anal. **21** (1984), 755–769.

5. T. F. Coleman and D. C. Sorensen, *A note on the computation of an orthonormal basis for the null space of a matrix*, Mathematical Programming **29** (1984), 234–242.

6. R. S. Dembo, S. C. Eisenstat, and T. Steihaug, *Inexact Newton methods*, SIAM J. Num. Anal. **19** (1982), 400–408.

7. J. E. Dennis and J. J. Moré, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Mathematics of Computation **28** (1974), 549–560.

8. ——, *Quasi-Newton methods, motivation and theory*, SIAM Review **19** (1977), 46–89.

9. R. Fontecilla, T. Steihaug, and R. A. Tapia, *A convergence theory for a class of quasi-Newton methods for constrained optimization*, SIAM J. Num. Anal. **24** (1987), 1133–1151.

10. P. Gill, W. Murray, M. Saunders, G. Stewart, and M. Wright, *Properties of a representation of a basis for the null space*, Mathematical Programming **33** (1985), 172–186.

11. J. Goodman, *Newton's method for constrained optimization*, Mathematical Programming **33** (1985), 162–171.

12. J. Nocedal and M. Overton, *Projected Hessian updating algorithms for nonlinearly constrained optimization*, SIAM J. Num. Anal. **22** (1985), 821–850.

13. J. Stoer and R. Tapia, *On the characterization of Q-superlinear convergence of quasi-Newton methods for constrained optimization*, Tech. Rep. 84-2, Dept. of Mathematical Sciences, Rice University (1984, revised 1986).

14. R. Tapia, *A stable approach to Newton's method for optimization problems with equality constraints*, Journal of Optimization Theory and Applications **14** (1974), 453–476.

CORNELL UNIVERSITY