

## The Convergence of the Ben-Israel Iteration for Nonlinear Least Squares Problems

By Paul T. Boggs

**Abstract.** Ben-Israel [1] proposed a method for the solution of the nonlinear least squares problem  $\min_{x \in D} \|F(x)\|_2$  where  $F: D \subset R^n \rightarrow R^m$ . This procedure takes the form  $x_{k+1} = x_k - F'(x_k)^+ F(x_k)$  where  $F'(x_k)^+$  denotes the Moore-Penrose generalized inverse of the Fréchet derivative of  $F$ . We give a general convergence theorem for the method based on Lyapunov stability theory for ordinary difference equations. In the case where there is a connected set of solution points, it is often of interest to determine the minimum norm least squares solution. We show that the Ben-Israel iteration has no predisposition toward the minimum norm solution, but that any limit point of the sequence generated by the Ben-Israel iteration is a least squares solution.

**I. Introduction.** The use of least squares solutions to systems of equations is an important and practical tool in many applications. Given a function  $F: D \subset R^n \rightarrow R^m$  where  $D$  is an open convex set, the nonlinear least squares problem is expressed as  $\min_{x \in D} \|F(x)\|$ , where  $\|\cdot\|$  here and henceforth denotes the  $l_2$  norm. Equivalently, if  $f_i(x)$  is the  $i$ th component of  $F$ , then the problem can be stated as  $\min_{x \in D} \Phi(x)$ , where  $\Phi = \frac{1}{2} \sum_{i=1}^m f_i^2(x)$ . If, as we shall assume,  $F$  is continuously Fréchet differentiable, then the minimum occurs where

$$(1.1) \quad \nabla \Phi(x) = F'(x)^T F(x) = 0.$$

In our discussion we shall always assume that there is at least one point in  $D$  satisfying (1.1).

With this assumption, any minimization procedure could be applied in an attempt to solve the problem. Several special purpose procedures, however, have been proposed which perform quite well. For example, if  $F'(x)$  has full rank in an open neighborhood of an isolated solution point, say  $x^*$ , then the Gauss-Newton iteration

$$(1.2) \quad x_{k+1} = x_k - t_k (F'(x_k)^T F'(x_k))^{-1} F'(x_k)^T F(x_k)$$

can be used. (Here  $t_k$  denotes the steplength.) If  $F'(x)$  is not of full rank, then the Levenberg-Marquardt [11] and [13] procedure is quite useful. This algorithm takes the form

$$(1.3) \quad x_{k+1} = x_k - t_k (\mu I + F'(x_k)^T F'(x_k))^{-1} F'(x_k)^T F(x_k),$$

---

Received August 7, 1974; revised June 19, 1975.

AMS (MOS) subject classifications (1970). Primary 65F20; Secondary 34D20.

Key words and phrases. Ben-Israel iteration, generalized inverses, nonlinear least squares, Lyapunov stability for difference equations.

Copyright © 1976, American Mathematical Society

where  $\mu$  is a positive constant. These algorithms have been studied extensively. For example, Brown and Dennis [4] have studied (1.2) and (1.3) in conjunction with finite-difference approximations to the derivative and have proposed further modifications of (1.3) to improve the performance. Also, Boggs and Dennis [3] obtain bounds on the final errors obtained if more general approximations are used. Golub and Pereyra [8] have proposed algorithms for nonlinear least squares problems whose variables separate.

The Ben-Israel iteration [1], with which we are primarily concerned, is designed for the case when  $F'(x)$  does not have full rank. We consider the following simple derivation of the iteration which will be useful later.

Let  $x_0$  be a given point. Then approximate  $F$  by a Taylor series through two terms to get

$$(1.4) \quad F(x) \approx F(x_0) + F'(x_0)(x - x_0)$$

and obtain the minimum norm solution vector  $(x - x_0)$  to the linear problem

$$\min_x \|F(x_0) + F'(x_0)(x - x_0)\|^2,$$

which is given by  $x - x_0 = -F'(x_0)^+ F(x_0)$  where  $F'(x_0)^+$  is the Moore-Penrose generalized inverse. (See Rao and Mitra [16] for a discussion of generalized inverses and linear least squares problems.) Thus the algorithm is given as

$$(1.5) \quad x_{k+1} = x_k - F'(x_k)^+ F(x_k).$$

Note that if  $F'$  has full rank, (1.5) reduces to the Gauss-Newton iteration (1.2).

Ben-Israel [1] gives numerical examples and a convergence result for this procedure, but the conditions for the theorem are somewhat restrictive and unnatural. In Boggs and Dennis [3], an analysis, based on classical stability theory from ordinary differential equations, is given; but here too the conditions given are somewhat restrictive. In the present work, we present very reasonable conditions on the function  $F$  which ensure convergence.

The Ben-Israel iteration and other generalized inverse methods have been used in practice to find least squares solutions. Some of these, along with computationally efficient modifications are discussed in Fletcher [6]. These methods are also sometimes used as portions of other algorithms. For example, Deuffhard [5] uses the Ben-Israel method in conjunction with an algorithm for solving multiple shooting equations. To fully understand these methods it is important to know first the fundamental characteristics of the underlying Ben-Israel iteration. The purpose of this paper is to supply such an analysis.

An important variation on the least squares problem arises when there is a connected set of points which solve the least squares problem, i.e., which satisfy (1.1). In this case, it is often of interest to find the minimum norm least squares solution. (If the set of points satisfying (1.1) is not connected, then the best we can hope for is a local minimum norm solution. Other procedures must then be used to find global solutions; we do not consider this problem here.) We investigate the application of the Ben-Israel iteration in this case and show that any limit point generated by it is a least

squares solution but not in general the minimum norm solution.

Our approach to this problem is to exploit the powerful Lyapunov stability theory for ordinary difference equations. In Boggs and Dennis [3], classical stability theory is used to obtain results on the application of nonlinear iterations where approximations are used. This theory rests on being able to solve the linearized problem at an isolated solution point  $x^*$ . The analysis breaks down if the linearized problem there is singular. The Lyapunov theory is powerful enough to handle the problem of an isolated singularity at an isolated solution point  $x^*$  and is extended (in Section 2) to the case of a connected set of solution points. (Related definitions and results are also given in Section 2.)

In Section 3 we apply the results of Section 2 to analyze the behavior of the Ben-Israel iteration.

**II. Lyapunov Stability.** The second method of Lyapunov [12] has been widely used for the study of the stability properties of systems of ordinary differential equations. (See, e.g. Sánchez [17] for an introduction or Yoshizawa [18] for a more complete account.) This theory has been extended to systems of ordinary difference equations and we refer to the paper of Ortega [14] for an account of the major results and related bibliography. (See also Hurt [10] and Hahn [9].) Ortega also nicely relates the concepts of stability and asymptotic stability to those of local convergence and attractiveness. The applications to date of this theory however, seem to be rather limited, applying, for example, to the stability of certain methods (such as Newton's method) under the influence of rounding errors. We propose to make more extensive use of the theory here and, as Ortega has done, to recommend the further exploitation of this powerful technique in the analysis of iterative methods.

For completeness, we give the definitions and state the relevant results. First, we give the material for the differential equation,

$$(2.1) \quad x' = -G(x), \quad x(0) = x_0$$

and then for the difference equation

$$(2.2) \quad x_{k+1} = x_k - t_k G(x_k), \quad x_0 \text{ given,}$$

where  $G: D \subset R^n \rightarrow R^n$ . (Assume throughout the remainder of the paper that  $G$  is a continuous function and that  $D$  is open and convex.) Clearly, (2.2) is just Euler's method applied to (2.1). We consider these forms since all iterative methods can be expressed in the form (2.2) and, therefore, (2.1) can be viewed as the related differential equation. This is useful since it is often easier to work first with the differential equation to obtain qualitative information on its behavior and then to use this information to analyze the iterative method. (Equation (2.1) is often referred to as the *continuous analogue* of (2.2). See Gavurin [7] and Boggs and Dennis [3] for further discussion of this connection.)

We begin by establishing the terminology.

*Definition 2.1.* A point  $x^*$  is

(a) *stable for (2.1)* if for any  $\epsilon > 0$  there is a  $\delta > 0$  such that if  $\|x_0 - x^*\| < \delta$ , then  $\|x(s) - x^*\| < \epsilon$  for  $s > 0$ , where  $x(s)$  is any solution to (2.1),

(b) *stable for (2.2)* if for any  $\epsilon > 0$  there is a  $\delta > 0$  such that if  $\|x_0 - x^*\| < \delta$ , then  $\|x_k - x^*\| < \epsilon$  for  $k > 0$ , where  $\{x_k\}$  is any solution to (2.2),

(c) *asymptotically stable for (2.1)* if it is stable for (2.1) and  $\lim_{s \rightarrow \infty} x(s) = x^*$ , and

(d) *asymptotically stable for (2.2)* if it is stable for (2.2) and  $\lim_{k \rightarrow \infty} x_k = x^*$ .

The idea of Lyapunov is to find a function which has certain properties on a solution curve of (2.1) which then forces  $x^*$  to be stable or asymptotically stable. We remark here that we will also need to include an assumption on  $G$  to ensure that a solution exists.

*Definition 2.2.* Let  $x(s)$  be any solution of (2.1). Let  $x^* \in D$  be such that  $G(x^*) = 0$ . Then a *Lyapunov function for (2.1) at  $x^*$*  is a continuously differentiable map  $V: D \subset R^n \rightarrow R^1$  such that

$$(2.3) \quad V(x) > 0, \quad x \neq x^*; \quad V(x^*) = 0$$

and

$$(2.4) \quad \frac{d}{ds} V(x(s)) \leq 0, \quad 0 \leq s < \infty.$$

Note that  $dV(x(s))/ds = V'(x(s))^T x'(s) = -V'(x(s))^T G(x)$ .

The major stability result is contained in the following theorem.

**THEOREM 2.3.** *Let  $V$  be a Lyapunov function for (2.1) at  $x^*$ . Then  $x^*$  is stable. If, in addition,*

$$(2.5) \quad \frac{d}{ds} V(x(s)) < 0, \quad 0 \leq s < \infty, \quad x(s) \neq x^*,$$

*then  $x^*$  is asymptotically stable.*

The corresponding results for difference equations differ only in the condition (2.4). We express the definition in terms of (2.2) in order to take into account the steplength. This formulation is somewhat different than the usual formulation, but will be useful in analyzing variable steplength algorithms.

*Definition 2.4.* Let  $\{x_k\}$  be a solution of (2.2). Let  $x^* \in D$  be such that  $G(x^*) = 0$ . Let  $\underline{t}$  and  $\bar{t}$  be constants such that  $0 < \underline{t} < \bar{t} < \infty$ . Then a *Lyapunov function for (2.2) at  $x^*$  for steplengths  $t_k \in [\underline{t}, \bar{t}]$*  is a continuously differentiable map  $V: D \subset R^n \rightarrow R^1$  such that (2.3) holds and

$$(2.6) \quad V(x - tG(x)) \leq V(x) \quad \text{for } x \in D \text{ and } t \in [\underline{t}, \bar{t}].$$

We remark that  $\bar{t}$  must of course always be chosen to ensure that  $x - tG(x) \in D$ .

**THEOREM 2.5.** *Let  $V$  be a Lyapunov function for (2.2) at  $x^*$ . Then  $x^*$  is stable. If, in addition*

$$V(x - tG(x)) < V(x) \quad \text{for } x \in D, \quad x \neq x^* \text{ and } t \in [\underline{t}, \bar{t}],$$

*then  $x^*$  is asymptotically stable.*

We remark here that it is sometimes difficult to find an appropriate Lyapunov function; but, as mentioned earlier, it is often easier to find a Lyapunov function for the continuous analogue (2.1) and then use the same function for (2.2). See Theorem 2.6.

As an example, consider steepest descent for minimizing a function  $f: D \subset R^n \rightarrow R^1$ . The iteration has the form

$$(2.7) \quad x_{k+1} = x_k - t_k \nabla f(x_k),$$

and the continuous analogue is

$$(2.8) \quad x' = -\nabla f(x).$$

Assume that  $f$  is continuously differentiable and that  $x^* \in D$  is the unique point such that  $\nabla f(x^*) = 0$ . Furthermore, suppose that  $f(x) > f(x^*)$  for all  $x \in D$  and  $x \neq x^*$ . Then  $V(x) = f(x) - f(x^*)$  is a Lyapunov function for (2.8):  $V(x)$  clearly satisfies (2.3) and

$$\frac{d}{ds} V(x(s)) = \nabla f(x)^T (-\nabla f(x)) = -\|\nabla f(x)\|^2 \leq 0$$

with equality only when  $x = x^*$ .

Rather than show that  $V(x) = f(x) - f(x^*)$  is also a Lyapunov function for (2.7) we prove a general theorem which will be of use in Section 3.

**THEOREM 2.6.** *Let  $V$  be a Lyapunov function for (2.1) at  $x^*$ . Assume  $V'$  is Lipschitz continuous with constant  $K$  on  $D$ . Suppose there is a constant  $c$  independent of  $x$  such that  $V'(x)^T G(x) \geq c\|G(x)\|^2$ . Then there are constants  $\underline{t}$  and  $\bar{t}$  such that  $V$  is a Lyapunov function for (2.2) at  $x^*$  for steplengths  $t \in [\underline{t}, \bar{t}]$ . Furthermore,  $\bar{t} < 2c/K$ .*

*Proof.* We need only show that (2.6) is satisfied. We have

$$\begin{aligned} V(x_{k+1}) - V(x_k) &= V(x_k - t_k G(x_k)) - V(x_k) \\ &= \{V(x_k - t_k G(x_k)) - V(x_k) + t_k V'(x_k)^T G(x_k)\} \\ &\quad + [V(x_k) - t_k V'(x_k)^T G(x_k)] - V(x_k). \end{aligned}$$

By the Lipschitz condition and by Ortega and Rheinboldt [15, Theorem 3, 2.12] the term in braces is bounded by  $\frac{1}{2}Kt_k^2\|G(x_k)\|^2$ . Therefore,

$$\begin{aligned} V(x_{k+1}) - V(x_k) &\leq -t_k V'(x_k)^T G(x_k) + \frac{1}{2}Kt_k^2\|G(x_k)\|^2 \\ &\leq [-t_k c + \frac{1}{2}Kt_k^2]\|G(x_k)\|^2, \end{aligned}$$

which is  $< 0$  if  $t_k c > \frac{1}{2}Kt_k^2$ . Choose  $\bar{t} < 2c/K$  and  $\underline{t}$  such that  $0 < \underline{t} < \bar{t} < 2c/K$ ; and therefore, for  $t \in [\underline{t}, \bar{t}]$  the result follows.

Returning to the steepest descent example, we see that  $G(x) = \nabla f(x)$  and  $V'(x)^T G(x) = \|G(x)\|^2$  so that  $c = 1$ . Thus, the steplengths are restricted to be in the interval  $[\underline{t}, 2/K]$ . This is a generalization of a result obtained in Boggs and Dennis

[3], but note that here the result depends on the choice of  $V$ .

In the introduction, we outlined the minimum norm least squares problem which we now state formally as

$$(2.9) \quad \min x^T x,$$

$$(2.10) \quad \text{subject to } F'(x)^T F(x) = 0.$$

Let  $Z = \{x: F'(x)^T F(x) = 0\}$  be a connected set.  $Z$  is the feasible region for the non-linear program (2.9)–(2.10) and we are interested in the “stability” of the set  $Z$ , i.e., we desire that if  $x_0$  is chosen close enough to  $Z$ , then the sequence generated by (2.2) should also remain close to  $Z$ . We first extend our definitions and then obtain the desired result. We shall use the standard definition of the distance from a point to a set; namely, if  $Y$  is any connected set then

$$\text{dist}(x, Y) \equiv \inf \{\|x - y\| : y \in Y\}.$$

*Definition 2.7.* A connected set  $Y$  is

- (a) *stable for (2.2)* if for any  $\epsilon > 0$  there is a  $\delta > 0$  such that if  $\text{dist}(x_0, Y) < \delta$ , then  $\text{dist}(x_k, Y) < \epsilon$  for all  $k > 0$ , where  $\{x_k\}$  is a solution to (2.2),
- (b) *asymptotically stable for (2.2)* if it is stable for (2.2) and  $\lim_{k \rightarrow \infty} \text{dist}(x_k, Y) = 0$ .

*Definition 2.8.* Let  $\{x_k\}$  be a solution of (2.2). Let  $Z$  be a connected set such that  $G(x) = 0$  for all  $x \in Z$ . Let  $\underline{t}$  and  $\bar{t}$  be constants such that  $0 < \underline{t} < \bar{t} < \infty$ . Then a *Lyapunov function for (2.2) on  $Z$*  for steplengths  $t_k \in [\underline{t}, \bar{t}]$  is a continuously differentiable map  $V: D \subset R^n \rightarrow R^1$  such that

$$(2.11) \quad V(x) > 0 \quad \text{for all } x \in D - Z, \quad V(x) = 0 \quad \text{for all } x \in Z,$$

and

$$(2.12) \quad V(x - tG(x)) \leq V(x) \quad \text{for all } x \in D \text{ and } t \in [\underline{t}, \bar{t}].$$

The proof of the next theorem follows closely that given in Ortega [14] for the proof of Theorem 2.5.

**THEOREM 2.9.** *Let  $Z \subset D$ ,  $\underline{t}$  and  $\bar{t}$  be as in Definition 2.8. Let  $V$  be a Lyapunov function for (2.2) on  $Z$ . Assume that  $V$  satisfies the condition: for every  $\epsilon > 0$  there is a  $\delta > 0$  such that  $V(x) > \epsilon$  if  $\text{dist}(x, Z) > \delta$ . Assume that  $G$  satisfies the condition: for every  $\epsilon > 0$  there is a  $\delta > 0$  such that if  $\text{dist}(x, Z) < \delta$ , then  $\|G(x)\| < \epsilon$ . Then  $Z$  is stable for (2.2). Furthermore, if*

$$(2.13) \quad V(x - tG(x)) < V(x) \quad \text{for all } x \in D, x \notin Z \text{ and } t \in [\underline{t}, \bar{t}],$$

*then  $Z$  is asymptotically stable for (2.2).*

*Proof.* Choose  $r$  such that  $\{x: \text{dist}(x, Z) \leq r\} \subset D$ . By the hypothesis on  $G$ , we can find an  $r_1 \leq r$  satisfying  $\text{dist}(x_{k+1}, Z) \leq r$  when  $\text{dist}(x_k, Z) \leq r_1$ . Then, choose any  $\eta \leq r_1$  and  $\gamma \in (0, \eta)$  in such a way that for any  $x$  satisfying  $\text{dist}(x, Z) \leq \gamma$  it follows that

$$V(x) \leq \psi(\eta) \equiv \inf \{V(y) : \eta \leq \text{dist}(y, Z) \leq r\} > 0.$$

Now suppose that there is an  $x_0$  such that  $\text{dist}(x_0, Z) < \gamma$  but for some  $k$ ,  $\text{dist}(x_{k+1}, Z) > \eta$ . Assume that this is the first such  $k$ . Then  $\text{dist}(x_i, Z) \leq \eta \leq r_1$ ,  $i = 1, \dots, k$ . Therefore, it follows that  $\text{dist}(x_{k+1}, Z) \leq r$  and  $V(x_{k+1})$  is well defined. Furthermore,  $V(x_{k+1}) \geq \psi(\eta)$ . But, by (2.12)

$$V(x_{k+1}) \leq V(x_k) \leq \dots \leq V(x_0) < \psi(\eta),$$

which is a contradiction; and hence, stability is proved.

Now assume (2.13) holds and let  $\hat{x}$  be any limit point of  $\{x_k\}$ . Let  $k_i$  be a sequence of indices such that  $x_{k_i} \rightarrow \hat{x}$  as  $i \rightarrow \infty$  and suppose that  $\hat{x} \notin Z$ . Then,  $r(x) = V(x - tG(x))/V(x)$  is well defined and continuous in an open neighborhood of  $\hat{x}$ . By (2.13),  $r(x) < 1$ . Thus, for  $\alpha \in (r(x), 1)$  there is a  $\gamma > 0$  such that  $r(x) \leq \alpha$ , when  $\|x - \hat{x}\| \leq \gamma$ . Thus, for  $k_i$  sufficiently large,

$$V(x_{k_i+1}) \leq \alpha V(x_{k_i}) \leq \alpha^2 V(x_{k_i-1}) \leq \dots \leq \alpha^i V(x_0).$$

Therefore,  $V(x_{k_i}) \rightarrow 0$  as  $i \rightarrow \infty$ ; and therefore,  $\hat{x} \in Z$  which completes the proof.

We complete this section by stating the analogous result to Theorem 2.6. The proof is the same; and hence, the details are omitted.

**THEOREM 2.10.** *Let  $V$  be a Lyapunov function for (2.1) on the set  $Y$  and assume that  $V'$  is Lipschitz continuous with constant  $K$  on  $D$ . Let  $c$  be a constant independent of  $x$  such that  $V'(x)^T G(x) \geq c \|G(x)\|^2$ . Then, there exist constants  $\underline{t}$  and  $\bar{t}$  so that  $V$  is also a Lyapunov function for (2.2) on the set  $Y$  when  $t_k \in [\underline{t}, \bar{t}]$ ,  $k > 0$ . Furthermore,  $\bar{t} < 2c/K$ .*

**III. The Ben-Israel Iteration.** In developing the analysis for the Ben-Israel iteration, we first examine the continuous analogue and then make use of Theorems 2.6 and 2.10. For convenience, we restate the method and the continuous analogue as

$$(3.1) \quad x_{k+1} = x_k - t_k F'(x_k)^+ F(x_k)$$

and

$$(3.2) \quad x' = -F'(x)^+ F(x),$$

where  $F: D \subset R^n \rightarrow R^m$ .

**LEMMA 3.1.** *Let  $F$  be continuously Fréchet differentiable on the open convex set  $D$ . Assume  $F'$  has constant rank  $r \leq \min(m, n)$  on  $D$ . Let  $x^*$  be the unique point in  $D$  satisfying  $F'(x^*)^T F(x^*) = 0$  and assume that  $\|F(x)\| > \|F(x^*)\|$  for all  $x \in D$ ,  $x \neq x^*$ . Then,*

$$(3.3) \quad V(x) = \frac{1}{2} \|F(x)\|^2 - \frac{1}{2} \|F(x^*)\|^2$$

is a Lyapunov function for (3.2) at  $x^*$ .

*Proof.* Clearly  $V(x) > 0$  for  $x \neq x^*$  and  $V(x^*) = 0$ . The condition that  $F'$  has constant rank ensures that  $F'(x)^+$  is continuous on  $D$  (see Golub and Pereyra [8]),

which implies that (3.2) has a solution. Now for  $x(s)$  any solution to (3.2),

$$\frac{d}{ds} V(x(s)) = - [F'(x)^T F(x)]^T F'(x)^+ F(x) = - F(x)^T F'(x) F'(x)^+ F(x).$$

First note that  $F'(x)F'(x)^+$  is the orthogonal projector on the column space of  $F'(x)$  and, thus, is positive semidefinite. Therefore,  $dV(x(s))/ds \leq 0$ .

Now since  $F'(x)^T F(x) \neq 0$  for  $x \neq x^*$ , it follows that  $F(x) \notin \text{Null}(F'(x)^T)$ . But  $\text{Null}(F'(x)^T) = (\text{Range } F'(x)^T)^\perp$ , and so  $F(x) \perp F'(x)F'(x)^+ F(x)$ ; and the proof is complete.

We may now prove the major Ben-Israel convergence theorem.

**THEOREM 3.2.** *Assume the conditions of Lemma 3.1 and that  $F'(x)^T F(x)$  is Lipschitz continuous with constant  $K$  on  $D$ . Then, there are constants  $\underline{t}$  and  $\bar{t}$  and a compact set  $D_0 \subset D$  such that  $x^* \in D_0$  and the sequence generated by (3.1) from any  $x_0 \in D_0$  with  $t_k \in [\underline{t}, \bar{t}]$  converges to  $x^*$ .*

*Proof.* We need to show that

$$V'(x)^T F'(x)^+ F(x) \geq c \|F'(x)^+ F(x)\|^2$$

for a constant  $c$  independent of  $x$  where  $V$  is given by (3.3). We have that  $V'(x) = F'(x)^T F(x)$ ; and hence,

$$\begin{aligned} V'(x)^T F'(x)^+ F(x) &= F^T(x) F'(x) F'(x)^+ F(x) \\ &= [F'(x)^+ F(x)]^T F'(x)^T F'(x) [F'(x)^+ F(x)]. \end{aligned}$$

From the proof of Lemma 3.1, we know that  $V'(x)^T F'(x)^+ F(x) > 0$  for all  $x \neq x^*$ . Therefore,  $F'(x)^+ F(x)$  is not in the null space of  $F'(x)^T F'(x)$  for  $x \neq x^*$ ; and we may conclude that  $V'(x)^T F'(x)^+ F(x) > \lambda_1(x) \|F'(x)^+ F(x)\|^2$ , where  $\lambda_1(x)$  is the smallest nonzero eigenvalue of  $F'(x)^T F'(x)$ . Now for some compact set  $D_0 \subset D$ ,  $\lambda_1(x)$  will be uniformly bounded away from zero by a number, say  $\bar{\lambda}$ , for all  $x \in D_0$ . Therefore, by Theorem 2.6,  $V$  is also a Lyapunov function for (3.1) for  $0 < \underline{t} < \bar{t} < 2\bar{\lambda}/K$ . We now apply Theorem 2.5 to complete the proof.

It has been pointed out by Professor J. J. Moré that the above proof could also be used to show that the Ben-Israel directions are gradient related in the sense of Ortega and Rheinboldt [15]. This connection has been further investigated and is reported in Boggs [2].

**THEOREM 3.3.** *Let  $F$  be continuously differentiable and  $F'(x)^T F(x)$  be Lipschitz continuous on the open convex set  $D$ . Let  $F'$  have constant rank  $r \leq \min(m, n)$  on  $D$ . Let  $D \supset Z = \{x: F'(x)^T F(x) = 0\}$  be bounded. Then there are constants  $\underline{t}$  and  $\bar{t}$  and a compact set  $D_0 \subset D$  such that  $Z \subset D_0$  and any limit point of the sequence generated by (3.1) from any  $x_0 \in D_0$  using steplengths  $t_k \in [\underline{t}, \bar{t}]$  is a member of  $Z$ .*

*Proof.* Let

$$V(x) = \frac{1}{2} [\|F(x)\|^2 - \|F(x^*)\|^2],$$

where  $x^*$  is the minimum norm solution. Since  $F'(x)^T F(x) = 0$  on  $Z$ , it follows that  $\|F(x)\|$  is constant on  $Z$ . Thus,  $V$  satisfies (2.11). The satisfaction of (2.12) follows



exactly as in the proof of Lemma 3.1; and hence, by Theorem 2.10,  $V(x)$  is a Lyapunov function for (3.2) on the set  $Z$ . The boundedness of  $Z$  implies the uniformity conditions on  $V$  and  $G$ ; and hence, the result follows by an application of Theorem 2.9.

It is of interest to know when a steplength of one may be used. This question has been considered by Brown and Dennis [4] and Boggs and Dennis [3] for the Gauss-Newton method, but no results were obtained for the Ben-Israel iteration. From Theorems 2.6 and 2.10 we know that  $\bar{t} < 2c/K$  and from the choice of  $V(x)$ , it follows that  $c = 1$  and  $K$  may be taken as the largest eigenvalue of  $(F'(x)^T F(x))'$  evaluated at  $x^*$ . Thus, if this eigenvalue is  $< 2$ , the steplengths may eventually be taken to be 1. This is exactly the same condition obtained in [3] and [4] for the Gauss-Newton method.

In order to determine the behavior of the Ben-Israel iteration, we consider its application to the following linear least squares problem:  $\min_x \|Ax - b\|$ , where

$$A = \begin{pmatrix} a & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$a \neq 0$ , and  $b$  is arbitrary. Clearly, the rank of  $A$  is one and the minimum norm solution is given by  $A^+ b$  which is

$$\begin{pmatrix} 1/a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} b_1/a \\ 0 \end{pmatrix}.$$

However, application of the Ben-Israel iteration from the initial guess  $x_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  yields

$$x_1 = x_0 - A^+(Ax_0 - b) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} b_1/a \\ 0 \end{pmatrix} = \begin{pmatrix} b_1/a \\ 1 \end{pmatrix}.$$

The next iteration shows that  $x_1$  is a fixed point and is a solution to the problem in the sense that  $A^T(Ax_1 - b) = 0$ . Note that for  $x_0 = (u, v)^T$ , the Ben-Israel iteration will always produce  $x_1 = (b_1/a, v)^T$  for any values of  $u$  and  $v$ . Thus, even in the linear case, the Ben-Israel iteration does not produce the minimum norm solution unless the initial vector is properly chosen on the correct manifold. In the nonlinear case, the manifold is also nonlinear; and it is, therefore, even harder to choose a correct  $x_0$ .

We propose a modification to the Ben-Israel iteration which, at least in the linear case, always converges to the minimum norm solution. Recalling the derivation given in Section 1, we see that the Ben-Israel iteration produces the minimum norm *correction* vector,  $x - x_0$  to the linearized problem. We propose to actually obtain the minimum norm *solution* to the linearized problem. That is, we write the linearized problem as

$$\min_x \|F'(x_0)x + [F'(x_0) - F'(x_0)x_0]\|^2$$

and obtain the minimum norm solution

$$x = -F'(x_0)^+ [F(x_0) - F'(x_0)x_0]$$

so that in general our algorithm is

$$(3.4) \quad x_{k+1} = F'(x_k)^+ F'(x_k)x_k - F'(x_k)^+ F(x_k).$$

In the linear case, we have the problem  $\min_x \|Ax - b\|$  for given  $A$  and  $b$ . Let  $x_0$  be any initial approximation. Then, from (3.4) we have

$$x_1 = A^+ Ax_0 - A^+(Ax_0 - b) = A^+ b,$$

so that the minimum norm solution is always obtained. The application of (3.4) in the nonlinear case is currently being investigated.

**IV. Acknowledgements.** The author would like to thank J. E. Dennis, Jr. and J. J. Moré of Cornell University for several helpful suggestions. He would also like to thank the referee for supplying the relatively simple proof of Lemma 3.1.

Mathematics Division  
United States Army Research Office  
P.O. Box 12211  
Research Triangle Park, North Carolina 27709

1. A. BEN-ISRAEL, "A Newton-Raphson method for the solution of systems of equations," *J. Math. Anal. Appl.*, v. 15, 1966, pp. 243–252. MR 34 #5273.
2. P. T. BOGGS, "On the use of Lyapunov theory for the analysis of nonlinear iterative methods," *Proc. of 1975 Conference on Information Sciences and Systems*, April 2–4, 1975.
3. P. T. BOGGS & J. E. DENNIS, JR., *A Stability Analysis for Perturbed Nonlinear Iterative Methods*, Cornell Univ. Computer Science Tech. Rep. TR 74–200; *Math. Comp.*, v. 30, 1976, pp. 199–215.
4. K. M. BROWN & J. E. DENNIS, JR., "Derivative free analogues of the Levenberg-Marquardt and Gauss algorithms for nonlinear least squares approximation," *Numer. Math.*, v. 18, 1971/72, pp. 289–297. MR 46 #2859.
5. P. DEUFLHARD, "A modified Newton method for the solution of ill-conditioned systems of nonlinear equations with application to multiple shooting," *Numer. Math.*, v. 22, 1974, pp. 289–315. MR 50 #3584.
6. R. FLETCHER, "Generalized inverse methods for the best least squares solution of systems of non-linear equations," *Comput. J.*, v. 10, 1967/68, pp. 392–399. MR 36 #4800.
7. M. K. GAVURIN, "Nonlinear functional equations and continuous analogs of iterative methods," *Izv. Vysš. Učebn. Zaved. Matematika*, v. 1958, no. 5(6), pp. 18–31; English transl., Technical Report 68–70, Univ. of Maryland, College Park, Md., 1968. MR 25 #1380.
8. G. H. GOLUB & V. PEREYRA, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, v. 10, 1973, pp. 413–432. MR 49 #1753.
9. W. HAHN, "Über die Anwendung der Methode von Ljapunov auf Differenzgleichungen," *Math. Ann.*, v. 136, 1958, pp. 430–441. MR 20 #5982.
10. J. HURT, "Some stability theorems for ordinary difference equations," *SIAM J. Numer. Anal.*, v. 4, 1967, pp. 582–596. MR 36 #4839.
11. K. LEVENBERG, "A method for the solution of certain non-linear problems in least squares," *Quart. Appl. Math.*, v. 2, 1944, pp. 164–168. MR 6, 52.
12. A. M. LYAPUNOV, "Problème générale de la stabilité du mouvement," Kharkov, 1892; French transl., *Ann. Fac. Sci. Univ. Toulouse*, v. (2) 9, 1907, pp. 203–474; reprint, *Ann. of Math. Studies*, no. 17, Princeton Univ. Press, Princeton, N. J., 1949. MR 9, 34.
13. D. W. MARQUARDT, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, v. 11, 1963, pp. 431–441. MR 27 #3040.
14. J. M. ORTEGA, "Stability of difference equations and convergence of iterative processes," *SIAM J. Numer. Anal.*, v. 10, 1973, pp. 268–282. MR 49 #4281.

15. J. M. ORTEGA & W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York and London, 1970. MR 42 #8686.
16. R. C. RAO & S. K. MITRA, *Generalized Inverse of Matrices and Its Applications*, Wiley, New York, 1971. MR 49 #2780.
17. D. A. SÁNCHEZ, *Ordinary Differential Equations and Stability Theory: An Introduction*, Freeman, San Francisco, Calif., 1968. MR 37 #3076.
18. T. YOSHIZAWA, *Stability Theory by Liapunov's Second Method*, Publ. Math. Soc. Japan, no. 9, Math. Soc. Japan, Tokyo, 1966. MR 34 #7896.