

# ALGORITHMS FOR EQUILIBRIA, GAMES & SYSTEMS OF NONLINEAR EQUATIONS

We consider the extension of the basic optimal decision problem to situations involving multiple decision makers pursuing their individual distinct objectives, in *competition* with each other. The policy adopted by one decision maker affects the policies of others. Such competitive situations arise among individuals, firms as well as nations. A solution to the problem is the **Nash equilibrium** strategy which ensures that no decision maker will benefit by deviating from this equilibrium. We introduce the **best replay algorithm** as the intuitive solution to the computation of equilibria and discuss relaxed variants with improved convergence properties.

We consider **Newton-type algorithms**. These aim to satisfy the optimality conditions for each player. These algorithms are quite general and applicable to a wide range of problems involving the **solution of nonlinear systems of equations**.

## 1. COMPUTATION OF EQUILIBRIA AND SOLUTION OF NONLINEAR EQUATIONS

### 1.1 Games and equilibria

For the  $i^{\text{th}}$  decision maker,  $i = 1, \dots, N$ , we are specifically concerned with the competitive situation of a Nash game in which the  $i^{\text{th}}$  decision maker seeks to optimize the  $i^{\text{th}}$  objective with respect to the  $i^{\text{th}}$  set of decision variables. We thus have the problem

$$\min_{Y, U^i} \left\{ f^i(Y, U^1, U^2, \dots, U^i, \dots, U^N) \mid g(Y, U^1, U^2, \dots, U^i, \dots, U^N) = 0 \right\} \quad (1.1)$$

where  $U^i$  is the vector of decision variables and  $f^i$  is the objective function of the  $i^{\text{th}}$  decision maker. The equality constraints represent the model of the underlying system which can be used to evaluate  $Y$ , given the values of  $U^1, U^2, \dots, U^i, \dots, U^N$ . Thus, we have a mapping similar to (1.4.1), given by

$$Y = \mathcal{Y}(U^1, U^2, \dots, U^i, \dots, U^N). \quad (1.2)$$

This mapping can be used to eliminate the output variables,  $Y$ , from the problem and express (1.1) as

$$\min_{U^i} \left\{ \mathcal{F}^i(U^1, U^2, \dots, U^i, \dots, U^N) \right\}, i = 1, \dots, N, \quad (1.3)$$

where

$$\mathcal{F}^i(U^1, U^2, \dots, U^i, \dots, U^N) = f^i(\mathcal{Y}(U^1, U^2, \dots, U^i, \dots, U^N), U^1, U^2, \dots, U^i, \dots, U^N).$$

For this problem, the Nash equilibrium is the point from which no decision maker would wish to deviate unilaterally since such an action would not improve the position of that decision maker. Thus, the Nash equilibrium  $U^i_*$ ,  $i = 1, \dots, N$ , is defined by

$$\mathcal{F}^i (U_*^1, U_*^2, \dots, U_*^i, \dots, U_*^N) \leq \mathcal{F}^i (U_*^1, U_*^2, \dots, U^i, \dots, U_*^N), \forall U^i, i = 1, \dots, N$$

(see e.g. Ho, 1970; Başar and Olsder, 1982). Hence,  $\mathcal{F}^i (U_*^1, U_*^2, \dots, U_*^i, \dots, U_*^N)$  is the unconstrained minimum of  $\mathcal{F}^i$ , with respect to  $U^i$ , and this is evaluated simultaneously for all  $i = 1, \dots, N$ . According to the necessary conditions of an optimum, this is equivalent to the solution of the simultaneous system of equations

$$\nabla_{U^i} \mathcal{F}^i (U^1, U^2, \dots, U^i, \dots, U^N) = 0, i = 1, \dots, N. \quad (1.4)$$

We note that the second order optimality also requires that the Hessian of  $\mathcal{F}^i(\cdot)$ , with respect to  $U^i$ , should be positive definite *at the solution*. This is ensured if each  $\mathcal{F}^i(\cdot)$  is convex with respect to  $U^i$  in a neighbourhood of the solution. The point,  $(U_*^1, U_*^2, \dots, U_*^i, \dots, U_*^N)$ , satisfying (1.4), is a minimum of  $\mathcal{F}^i$ , with respect to  $U^i$ , and thus a Nash solution. Hence, at the Nash equilibrium, the objective function of the  $i^{\text{th}}$  decision maker satisfies

$$\langle v, \nabla_{U^i}^2 \mathcal{F}^i (U_*^1, U_*^2, \dots, U_*^i, \dots, U_*^N) v \rangle > 0, \forall v \neq 0, i = 1, \dots, N. \quad (1.5)$$

Clearly, for an algorithm solving (1.4), this is a difficult condition to ensure. It is therefore desirable that all points satisfying (1.4) also satisfy (1.5). In other words that  $\mathcal{F}^i(\cdot)$  is globally convex with respect to  $U^i$ .

**Algorithm 1.1 [Best replay]**

**Step 0:** Given  $U_0^1, U_0^2$ , macheps, set  $k = 0$ .

**Step 1:** Compute

$$U_{k+1}^1 = \arg \min_{U^1} \left\{ \mathcal{F}^1 (U^1, U_k^2) \right\}$$

and

$$U_{k+1}^2 = \arg \min_{U^2} \left\{ \mathcal{F}^2 (U_k^1, U^2) \right\}$$

**Step 2:** Equilibrium check: if

$$\left\| \begin{bmatrix} U_{k+1}^1 \\ U_{k+1}^2 \end{bmatrix} - \begin{bmatrix} U_k^1 \\ U_k^2 \end{bmatrix} \right\| \leq \text{macheps}$$

**stop.** Otherwise set  $k = k + 1$ , go to **Step 1**.

Where macheps is a tolerance value that depends on the accuracy to which each optimisation problem can be solved in practice. This is essentially a **Jacobi-type algorithm** with each player optimising given the previous iteration for the other player (see Ortega and Rheinbold, 1970).

The application of a **Gauss-Seidel-type procedure** (Ortega and Rheinbold, 1970) leads to **Step 1** being replaced by

$$U_{k+1}^1 = \arg \min_{U^1} \left\{ \mathcal{F}^1 (U^1, U_k^2) \right\}$$

and

$$U_{k+1}^2 = \arg \min_{U^2} \left\{ \mathcal{F}^2 (U_{k+1}^1, U^2) \right\}.$$

If the above procedures converge, then it can be verified that they converge to the solution. However, convergence may be slow and convergence itself cannot be always assured.

### 1.2 Solution of Systems of Equations

The remaining discussion is concerned with the solution of a system of nonlinear equations arising from (1.4) (see Rosen, 1965).

## 2. NEWTON-TYPE ALGORITHMS

Let the equilibrium condition for the **Nash strategy** (1.4) be written as

$$\boxed{\mathcal{W}(y) \equiv \nabla_{U^i} \mathcal{F}^i (U^1, U^2, \dots, U^i, \dots, U^N) = 0; \quad i = 1, \dots, N.} \quad (2.1)$$

where

$$y \equiv \begin{bmatrix} U^1 \\ \vdots \\ U^i \\ \vdots \\ U^N \end{bmatrix}$$

$U^1, U^2, \dots, U^i, \dots, U^N$ . Hence, let  $y \in \mathbb{R}^n$  and  $\mathcal{W} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . In this section, we consider an algorithm for solving the system of nonlinear equations  $\mathcal{W}(y) = 0$ .

**Definition.** The system evaluated at point  $y_j$  is denoted by  $\mathcal{W}_j = \mathcal{W}(y_j)$  and similarly, the Jacobian  $\nabla \mathcal{W}_j = \nabla \mathcal{W}(y_j)$ .

The direction of search  $d_j$ , given  $y_j$ , is determined by

$$\boxed{\nabla \hat{\mathcal{W}}_j d_j = - \mathcal{W}_j} \quad (2.2)$$

where  $\nabla \hat{\mathcal{W}}_j$  is an approximation to  $\nabla \mathcal{W}_j$ , introduced below, with periodic reevaluation of the exact  $\nabla \mathcal{W}_j$  at intervals within the algorithm. The algorithm using the **exact value**  $\nabla \mathcal{W}_j$  throughout is known as the **Newton algorithm**. The algorithm using the approximation  $\nabla \hat{\mathcal{W}}_j$  below is the **quasi-Newton algorithm**. The quasi-Newton algorithm then proceeds along  $d_j$  such that

$$\boxed{y_{j+1} = y_j + \tau_j d_j} \quad (2.3)$$

with  $\tau_j \in [0, 1]$  given by the smallest value of  $k = 0, 1, 2, 3, \dots$ ,  $\tau_j = (\pi)^k$ ,  $\pi \in (0, 1)$ , satisfying

$$\boxed{\frac{1}{2} \|\mathcal{W}_{j+1}\|_2^2 - \frac{1}{2} \|\mathcal{W}_j\|_2^2 \leq \tau_j \mu < [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j >, \quad \mu \in (0, \frac{1}{2}]} \quad (2.4)$$

**Definition.** The algorithms usually compute a selected number of columns of  $\nabla \mathcal{W}_j$  during an iteration. We use  $\nabla \hat{\mathcal{W}}_j$  to denote this and the Broyden update approximation to  $\nabla \mathcal{W}_j$

$$\nabla \hat{\mathcal{W}}_j = \nabla \hat{\mathcal{W}}_{j-1} + \frac{[\mathcal{W}_j - \mathcal{W}_{j-1} - \nabla \hat{\mathcal{W}}_{j-1} (y_j - y_{j-1})] (y_j - y_{j-1})^T}{\langle (y_j - y_{j-1}), (y_j - y_{j-1}) \rangle}. \quad (2.5)$$

<b>Advantages &amp; disadvantages of Newton's method</b>	
<i>Advantages</i>	
1.	fast convergence (see example above and theory below) from good starting points if $\nabla \mathcal{W}$ is nonsingular
2.	exact solution in one iteration for linear $\mathcal{W}$
<i>Disadvantages</i>	
1.	not globally convergent from arbitrary starting points so $\tau$ required to dampen the steplength
2.	requires 'exact' $\nabla \mathcal{W}$ evaluation (in contrast to $\nabla \hat{\mathcal{W}}$ ) at each iteration
3.	each iteration requires the solution of a linear system of equations that may be singular (due to singular $\nabla \mathcal{W}$ )

**EXERCISE:** Write a pseudo code for the full quasi-Newton algorithm above.

**EXAMPLE 1:** Consider the original Newton algorithm (i.e. using the exact value  $\nabla \mathcal{W}(y)$  and the choice  $\tau_j = 1$  always) as applied to

$$\mathcal{W}(y) = \begin{bmatrix} y^1 + y^2 - 3 \\ (y^1)^2 + (y^2)^2 - 9 \end{bmatrix} = 0.$$

There are two solutions to this problem  $(y^1, y^2) = (3, 0)$  or  $(0, 3)$ . Let the algorithm start with the initial estimate  $y_0 = (1, 5)$ . The first two iterations of Newton's algorithm are

$$\nabla \mathcal{W}(y_0) d_0 = - \mathcal{W}(y_0): \begin{bmatrix} 1 & 1 \\ 2 & 10 \end{bmatrix} d_0 = - \begin{bmatrix} 3 \\ 17 \end{bmatrix} \Rightarrow d_0 = - \begin{bmatrix} \frac{13}{8} \\ \frac{11}{8} \end{bmatrix}$$

$$y_1 = y_0 + d_0 = \begin{bmatrix} - .625 \\ 3.625 \end{bmatrix}$$

$$\nabla \mathcal{W}(y_1) d_1 = - \mathcal{W}(y_1): \begin{bmatrix} 1 & 1 \\ -\frac{5}{4} & \frac{29}{4} \end{bmatrix} d_1 = - \begin{bmatrix} 0 \\ \frac{145}{32} \end{bmatrix} \Rightarrow d_1 = \begin{bmatrix} \frac{145}{272} \\ -\frac{145}{272} \end{bmatrix}$$

$$y_2 = y_1 + d_1 \cong \begin{bmatrix} - .092 \\ 3.092 \end{bmatrix}.$$

Thus the algorithm is working well in this case.

**EXAMPLE 2:** Consider again the original Newton algorithm as applied to

$$\mathcal{W}(y) = \begin{bmatrix} e^{y^1} - 1 \\ e^{y^2} - 1 \end{bmatrix} = 0.$$

The solution being  $(y^1, y^2) = (0, 0)$ . Let the algorithm start with the initial estimate  $y_0 = (10, 10)$ , then we have

$$y_1 = \begin{bmatrix} -11 + e^{10} \\ -11 + e^{10} \end{bmatrix} \cong \begin{bmatrix} 2.2 \times 10^4 \\ 2.2 \times 10^4 \end{bmatrix}$$

which is not a very good step.

**EXAMPLE 3:** Consider again the original Newton algorithm as applied to

$$\mathcal{W}(y) = \begin{bmatrix} (y^1)^2 + (y^2)^2 - 2 \\ e^{y^1-1} + (y^2)^3 - 2 \end{bmatrix} = 0.$$

which is solved by  $(y^1, y^2) = (1, 1)$ . Let  $y_0 = (2, 0.5)$

$$\nabla \mathcal{W}(y_0) d_0 = -\mathcal{W}(y_0): \begin{bmatrix} 4 & 1 \\ e & 0.75 \end{bmatrix} d_0 = -\begin{bmatrix} 2.25 \\ 0.843 \end{bmatrix} \Rightarrow d_0 \cong \begin{bmatrix} -3.00 \\ 9.74 \end{bmatrix}$$

$$y_1 = y_0 + d_0 \cong \begin{bmatrix} -1.00 \\ 10.24 \end{bmatrix} \Rightarrow \mathcal{W}(y_1) = \begin{bmatrix} 104 \\ 1071 \end{bmatrix}$$

so the choice  $\tau_0 = 1$  is unsatisfactory. Trying  $\tau_0 = .1$ , we have

$$y_1 = y_0 + .1 d_0 \cong \begin{bmatrix} 1.70 \\ 1.47 \end{bmatrix} \Rightarrow \mathcal{W}(y_1) = \begin{bmatrix} 3.06 \\ 3.21 \end{bmatrix}$$

which is still unsatisfactory. We reduce the stepsize further to  $\tau_0 = .05$  which is still unsatisfactory

$$y_1 = y_0 + .05 d_0 \cong \begin{bmatrix} 1.85 \\ 0.97 \end{bmatrix} \Rightarrow \mathcal{W}(y_1) = \begin{bmatrix} 2.40 \\ 1.30 \end{bmatrix}.$$

We reduce it further to  $\tau_0 = .0116$

$$y_1 = y_0 + .0116 d_0 \cong \begin{bmatrix} 1.965 \\ 0.613 \end{bmatrix} \Rightarrow \mathcal{W}(y_1) = \begin{bmatrix} 2.238 \\ 0.856 \end{bmatrix}.$$

This point is satisfactory as it satisfies (2.4) (say with  $\mu = 10^{-4}$ )

$$\frac{1}{2} \|\mathcal{W}(y_1)\|_2^2 \cong 2.87 < 2.89 \cong \frac{1}{2} \|\mathcal{W}(y_0)\|_2^2 + (.0116)(10^{-4}) < [\nabla \mathcal{W}_0]^T \mathcal{W}_0, d_0 > .$$

Note: the gradual reduction of  $\tau_0$  is for illustration only. We could have just tried the first three terms of the series  $\tau_j = (0.1)^k$ ,  $k = 0, 1, 2$ . The last one,  $\tau_j = .01$ , would have also satisfied (2.4) (verify).

With the value  $y_1$  just determined, we proceed further to the next iteration with  $\tau_j = (0.1)^0 = 1$

$$\nabla \mathcal{W}(y_1) d_1 = -\mathcal{W}(y_1) \Rightarrow d_1 \cong \begin{bmatrix} -1.22 \\ 2.07 \end{bmatrix}$$

$$y_2 = y_1 + d_1 \cong \begin{bmatrix} 0.70 \\ 2.68 \end{bmatrix} \Rightarrow \mathcal{W}(y_2) \cong \begin{bmatrix} 5.77 \\ 18.13 \end{bmatrix}$$

which is again unsatisfactory so we try  $\tau_j = (0.1)^1 = 0.1$

$$y_2 = y_1 + 0.1 d_1 \cong \begin{bmatrix} 1.84 \\ 0.820 \end{bmatrix} \Rightarrow \mathcal{W}(y_2) \cong \begin{bmatrix} 2.07 \\ 0.876 \end{bmatrix}$$

and this stepsize is accepted since

$$\frac{1}{2} \|\mathcal{W}(y_2)\|_2^2 \cong 2.53 < 2.87 \cong \frac{1}{2} \|\mathcal{W}(y_1)\|_2^2 + (.1)(10^{-4}) < [\nabla \mathcal{W}_1]^T \mathcal{W}_1, d_1 > .$$

At the next iteration, we have

$$\nabla \mathcal{W}(y_2) d_2 = -\mathcal{W}(y_2) \Rightarrow d_2 \cong \begin{bmatrix} -0.0756 \\ 0.0437 \end{bmatrix}$$

$$y_3 = y_2 + d_2 \cong \begin{bmatrix} 1.088 \\ 1.257 \end{bmatrix} \Rightarrow \mathcal{W}(y_3) \cong \begin{bmatrix} 0.762 \\ 1.077 \end{bmatrix}$$

so the Newton step performs very well and from this iteration on, the algorithm converges to the solution (1, 1) quadratically.

**EXAMPLE 4:** Consider the use of the approximation formula (2.5) to solve the problem in Example 3

$$\mathcal{W}(y) = \begin{bmatrix} (y^1)^2 + (y^2)^2 - 2 \\ e^{y^1-1} + (y^2)^3 - 2 \end{bmatrix} = 0.$$

which is solved by  $(y^1, y^2) = (1, 1)$ . Let  $y_0 = (1.5, 2)$ . We assume  $\tau_j = 1$  throughout. It turns out that the algorithm using the approximation  $\nabla \hat{\mathcal{W}}$  converges to the solution in 10 iterations with

$$\nabla \hat{\mathcal{W}}_{10} = \begin{bmatrix} 1.999137 & 2.021829 \\ 0.9995643 & 3.011004 \end{bmatrix} \text{ while } \nabla \mathcal{W}(1,1) = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$$

and the Newton algorithm (using  $\tau_j = 1$  and exact value  $\nabla \mathcal{W}$  throughout) converges in 6 iterations. In this case,  $\nabla \hat{\mathcal{W}}_{10}$  closely approximates the value at the solution  $\nabla \mathcal{W}(1,1)$ . This may not happen as often as one would like.

**EXAMPLE 5:** Consider the use of the approximation formula (2.5) to solve the problem in Example 1.

$$\mathcal{W}(y) = \begin{bmatrix} y^1 + y^2 - 3 \\ (y^1)^2 + (y^2)^2 - 9 \end{bmatrix} = 0.$$

and the two solutions to this problem are  $(y^1, y^2) = (3, 0)$  or  $(0, 3)$ . It can be shown (see Dennis and Schnabel, 1983, Lemma 8.2.7) that for this example

$$\lim_{j \rightarrow \infty} \nabla \hat{\mathcal{W}}_j = \begin{bmatrix} 1 & 1 \\ 1.5 & 7.5 \end{bmatrix} \quad \text{when} \quad \nabla \mathcal{W}(0, 3) = \begin{bmatrix} 1 & 1 \\ 0 & 6 \end{bmatrix}$$

The algorithm using the approximation  $\nabla \hat{\mathcal{W}}$ , however, does converge to the solution as well as the Newton algorithm using the exact value  $\nabla \mathcal{W}(y)$ . Indeed, it can be verified that the former converges in 6 and the latter in 5 iterations.

To investigate how the former algorithm proceeds, let the initial value to start the algorithm be  $y_0 = (1, 5)$ . Then

$$\mathcal{W}(y_0) = \begin{bmatrix} 3 \\ 17 \end{bmatrix}; \quad \nabla \hat{\mathcal{W}}_0 = \nabla \mathcal{W}(y_0) = \begin{bmatrix} 1 & 1 \\ 2 & 10 \end{bmatrix}$$

$$\nabla \hat{\mathcal{W}}_0 d_0 = - \mathcal{W}(y_0) \quad \Rightarrow \quad d_0 = \begin{bmatrix} -1.625 \\ -1.375 \end{bmatrix}$$

$$y_1 = y_0 + d_0 = \begin{bmatrix} -0.625 \\ 3.625 \end{bmatrix} \quad \Rightarrow \quad \mathcal{W}(y_1) = \begin{bmatrix} 0 \\ 4.53125 \end{bmatrix}$$

$$\nabla \hat{\mathcal{W}}_1 = \nabla \hat{\mathcal{W}}_0 + \begin{bmatrix} 0 & 0 \\ -1.625 & -1.375 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0.375 & 8.625 \end{bmatrix}$$

**EXERCISE:** Confirm that  $\mathcal{W}(y_1) = \mathcal{W}(y_0) + \nabla \hat{\mathcal{W}}_1 d_0$  (hence  $\nabla \hat{\mathcal{W}}_1$  is a valid linear approximation between  $\mathcal{W}(y_1)$  and  $\mathcal{W}(y_0)$ ).

We note that

$$\nabla \mathcal{W}(y_1) = \begin{bmatrix} 1 & 1 \\ -1.25 & 7.25 \end{bmatrix}$$

so that  $\nabla \hat{\mathcal{W}}_1$  is not a good approximation to  $\nabla \mathcal{W}(y_1)$ . At the next iteration,

$$\nabla \hat{\mathcal{W}}_1 d_1 = - \mathcal{W}(y_1) \quad \Rightarrow \quad d_1 \cong \begin{bmatrix} 0.549 \\ -0.549 \end{bmatrix}$$

$$y_2 = y_1 + d_1 \cong \begin{bmatrix} -0.076 \\ 3.076 \end{bmatrix} \quad \Rightarrow \quad \mathcal{W}(y_2) = \begin{bmatrix} 0 \\ 0.466 \end{bmatrix}$$

$$\nabla \hat{\mathcal{W}}_2 \cong \begin{bmatrix} 1 & 1 \\ -0.799 & 8.201 \end{bmatrix} \quad \text{whereas} \quad \nabla \mathcal{W}(y_2) \cong \begin{bmatrix} 1 & 1 \\ -0.152 & 6.152 \end{bmatrix}$$

so the approximate  $\nabla \hat{\mathcal{W}}_2$  is again not really close to the actual value. This example illustrates that, **when the approximate algorithm, based on using  $\nabla \hat{\mathcal{W}}_j$ , converges almost as fast as the Newton algorithm, one cannot assume that the final approximation  $\nabla \hat{\mathcal{W}}$  will approximate the actual  $\nabla \mathcal{W}$  at the solution, although often it does.**

### 3. CONVERGENCE OF NEWTON-TYPE ALGORITHMS

In this section, we consider the convergence properties of the basic algorithm. We establish global convergence with relaxed descent directions and relate the achievement of unit stepsizes to the difference of the Jacobian approximation and the actual Jacobian at the solution. We also discuss the Q-superlinear convergence rate.

The global convergence of the algorithm mainly requires that  $d_j$  is a descent direction for the sum of squares merit function such that

$$\langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j \rangle \leq -m \|d_j\|^2 \quad (3.1, a)$$

for some  $m > 0$ . This is clearly satisfied by the steepest descent and, with a nonsingular Jacobian, the Newton directions. The quasi-Newton direction satisfies

$$\langle [\nabla \hat{\mathcal{W}}_j]^T \mathcal{W}_j, d_j \rangle \leq -m \|d_j\|^2. \quad (3.1, b)$$

The stepsize is adjusted to ensure the inequality

$$\boxed{\frac{1}{2} \|\mathcal{W}_{j+1}\|_2^2 - \frac{1}{2} \|\mathcal{W}_j\|_2^2 \leq \tau_j \mu \langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j \rangle} \quad (3.2)$$

where  $\langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j \rangle \leq 0$ . If, even for small stepsizes, the Newton or the quasi-Newton direction (2.2) does not satisfy (3.2), then the steepest descent direction  $d_j = -[\nabla \mathcal{W}_j]^T \mathcal{W}_j$  or even  $-d_j$  may be tried (Li, 1989).

In the algorithms of section 2, an approximate Jacobian is used. If (3.2) is not satisfied for  $\tau_j=1$ , with this approximation, then the algorithm computes the Jacobian numerically before attempting to reduce the stepsize in order to satisfy the stepsize criterion (3.2).

#### Assumption 3.1

1. There exists a solution,  $y_*$ , to  $\mathcal{W}(y) = 0$ .
2.  $\mathcal{W} \in C^1(\mathbb{R}^n)$ ;
3. The direction  $d_j$  satisfies (3.1).

Assumption 2 is required to ensure that the algorithm is well defined. We also require that  $\nabla \mathcal{W}$  is Lipschitz continuous to establish the monotonic decrease of the sequence  $\{\|\mathcal{W}_j\|_2^2\}$  discussed in Theorem 3.1 below and also to establish global convergence. The monotonic decrease may also be obtained without Lipschitz continuity if it is assumed that  $\|\mathcal{W}_j\| = \phi \|d_j\|$ , for some  $\phi \in (0, \infty)$ . The latter is satisfied both by the steepest descent with a nonsingular Jacobian approximation and with the Newton and quasi-Newton directions.

In Theorems 3.3 - 3.5, we further require  $\mathcal{W} \in \mathbb{C}^2(\mathbb{R}^n)$  to establish the achievement of unit stepsizes (i.e.  $\tau_j = 1$ ). The descent condition in Assumption 3 is satisfied by the Newton and steepest descent directions. For the quasi-Newton direction, the algorithm ensures satisfaction by refining the Jacobian approximation. If the descent condition is not satisfied, even after a full numerical evaluation of the Jacobian, with reduced  $\delta$ , in **Step 2**, then the algorithms fail. In Theorem 3.1, we have allowed for a general descent direction including the Newton, quasi-Newton or steepest descent steps. For example, one reasonable way of establishing the quasi-Newton case is by assuming  $\|\mathcal{W}_j\| = \phi \|\mathbf{d}_j\|$  as in Theorem 3.1, ii.

**Remark. The bound  $\mu \in (0, 1]$**

In Theorems 3.1 - 3.2, the bound  $\mu \in (0, 1]$  is sufficient. However, we have adopted the tighter bound  $\mu \leq \frac{1}{2}$  since the latter is required to establish the convergence to unit stepsizes.  $\square$

**Theorem 3.1 [Monotonic Decrease]**

Let Assumptions 3.1 be satisfied and furthermore let either

- (i)  $\nabla \mathcal{W}$  be Lipschitz continuous:  $\|\nabla \mathcal{W}(\mathbf{y}) - \nabla \mathcal{W}(\mathbf{z})\| \leq \ell \|\mathbf{y} - \mathbf{z}\|$ , for  $\ell > 0$ ; or,
- (ii)  $\mathbf{d}_j$  satisfies  $\|\mathcal{W}_j\| = \phi \|\mathbf{d}_j\|$ , for some  $\phi \in (0, \infty)$ .
- (iii)  $\mathbf{d}_j$  satisfies (3.1).

Then, the stepsize computed in the N-SCE and N-B-SCE algorithms is such that  $\tau_j \in (0, 1]$  and hence the sequence  $\{\mathbf{y}_j\}$  generates a corresponding **monotonically decreasing sequence**  $\{\|\mathcal{W}_j\|_2^2\}$ .

**Proof**

Using the first order expansion of  $\|\mathcal{W}\|_2^2$  we have

$$\frac{1}{2} \|\mathcal{W}_{j+1}\|_2^2 - \frac{1}{2} \|\mathcal{W}_j\|_2^2 = \tau_j \int_0^1 \langle [\nabla \mathcal{W}(\mathbf{y}(t))]^T \{\mathcal{W}(\mathbf{y}(t))\}, \mathbf{d}_j \rangle dt.$$

Furthermore, the relationships below arise from assumptions and definitions above:

$$\int_0^1 \langle \nabla \mathcal{W}_j^T \mathcal{W}_j - \nabla \mathcal{W}_j^T \mathcal{W}_j + \nabla \mathcal{W}(\mathbf{y}(t))^T \mathcal{W}_j - \nabla \mathcal{W}(\mathbf{y}(t))^T \mathcal{W}_j, \mathbf{d}_j \rangle dt = 0$$

$$\int_0^1 \langle \nabla \mathcal{W}_j^T \mathcal{W}_j, \mathbf{d}_j \rangle dt = \langle \nabla \mathcal{W}_j^T \mathcal{W}_j, \mathbf{d}_j \rangle$$

$$\mathbf{y}(t) = \mathbf{y}_j + t \tau_j \mathbf{d}_j \text{ and Lipschitz continuity } \Rightarrow$$

$$\|\nabla \mathcal{W}(\mathbf{y}(t)) - \nabla \mathcal{W}_j\| \leq \ell \|\mathbf{y}(t) - \mathbf{y}_j\| = t \tau_j \ell \|\mathbf{d}_j\|$$

$$\text{Assumption 3 \& (3.1)} \Rightarrow \langle \nabla \mathcal{W}_j^T \mathcal{W}_j, \mathbf{d}_j \rangle \leq -m \|\mathbf{d}_j\|^2$$

$$\frac{1}{2} \|\mathcal{W}_{j+1}\|_2^2 - \frac{1}{2} \|\mathcal{W}_j\|_2^2 = \tau_j \langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle$$

$$+ \tau_j \int_0^1 \langle [\nabla \mathcal{W}(\mathbf{y}(t))]^T \{\mathcal{W}(\mathbf{y}(t)) - \mathcal{W}_j\} + \{[\nabla \mathcal{W}(\mathbf{y}(t))]^T - [\nabla \mathcal{W}_j]^T\} \mathcal{W}_j, \mathbf{d}_j \rangle dt \tag{3.3}$$

where  $y(t) = y_j + t \tau_j d_j$ . For (i), given the Lipschitz continuity and (3.1, a) we have

$$\frac{1}{2} \|\mathcal{W}_{j+1}\|_2^2 - \frac{1}{2} \|\mathcal{W}_j\|_2^2 \leq \tau_j \langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j \rangle \left[ 1 - \frac{\tau_j}{2m} \left\{ \psi^2 + \ell \|\mathcal{W}_j\| \right\} \right] \quad (3.4)$$

where  $\|\nabla \mathcal{W}\| \leq \psi$ . The scalar  $\mu \in (0, 1)$  in the stepsize strategy (3.2) clearly determines  $\tau_j$  such that  $\mu \leq 1 - \frac{\tau_j}{2m} \left\{ \psi^2 + \ell \|\mathcal{W}_j\| \right\} \leq \frac{1}{2}$ . By (3.1), there exists a  $\tau_j \in [0, 1]$  satisfying the inequalities (3.2) and (3.3). Suppose  $\tau^0$  is the largest  $\tau \in [0, 1]$  satisfying these inequalities. Thus, all  $\tau \leq \tau^0$  also satisfy these conditions and that the stepsize strategy selects a  $\tau_j \in [\bar{\tau}\tau^0, \tau^0]$ , where  $\bar{\tau} \in (0, 1)$  is input in **Step 0**. By (3.1), it follows that  $\{\|\mathcal{W}_j\|_2^2\}$  is a monotonically decreasing sequence. For (ii), we can use (3.3) to derive a relationship similar to (3.4) by invoking  $\|\mathcal{W}_j\| = \phi \|d_j\|$ .  $\square$

We discuss the global convergence of the basic Newton and quasi-Newton algorithms.

**\*\*YOU ARE NOT RESPONSIBLE FOR ANY PROOFS BELOW IN THIS SET\*\*.**

**Lemma 3.1**

Let the assumptions of Theorem 3.1 be satisfied. We then have

$$\lim_{j \rightarrow \infty} \langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j \rangle = 0. \quad (3.5)$$

**Proof**

The level set

$$\mathbb{F} = \left\{ y \in \mathbb{R}^n \mid \|\mathcal{W}(y)\|_2^2 \leq \|\mathcal{W}(y_{j_0})\|_2^2 \right\}$$

is bounded for some  $j_0 \geq 0$ . We state the proof for the Lipschitz continuous case (i) in Theorem 3.1. Case (ii) can also be similarly established using the compactness of  $\mathbb{F}$ . Given  $\mu \in (0, \frac{1}{2})$ , by (3.4), the choice

$$\tau_0 = \min \left\{ 1, \frac{(1-\mu)}{\frac{1}{2m}(\psi^2 + \ell \|\mathcal{W}_j\|)} \right\}$$

always satisfies the stepsize strategy (3.2). Clearly,  $\tau_j$ , chosen in the algorithms is in the range  $\tau_j \in [\bar{\tau}\tau_0, \tau_0]$  and thereby also satisfies (3.2). As  $\mathcal{W}$  is Lipschitz continuous and  $\mathbb{F}$  is compact, there exists a scalar  $M < \infty$  and  $\|\mathcal{W}_j\| < M$ . Thus, as  $\ell \in (0, \infty)$  and  $m > 0$ , we have  $\tau_j \geq \hat{\tau} > 0, \forall j$ , for some  $\hat{\tau}$  and stepsize strategy (3.2). The boundedness of  $\|\mathcal{W}(y)\|_2^2$  on  $\mathbb{F}$  and  $\langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j \rangle \leq 0$  imply

$$0 \leq \mu \sum_j \tau_j \langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, d_j \rangle \leq \sum_j \left( \|\mathcal{W}_j\|_2^2 - \|\mathcal{W}_{j+1}\|_2^2 \right) < \infty$$

which yields (3.5).  $\square$

**Lemma 3.2**

Inequality (3.1) and Lemma 3.1 imply  $\lim_{j \rightarrow \infty} \|d_j\| = 0$ .

**Proof**

The result follows from (3.1) and (3.5). □

We can hence show the global convergence for any algorithm satisfying (3.1) as well as  $\|W_j\| = \phi \|d_j\|$ , for some  $\phi \in (0, \infty)$ . As mentioned earlier, this is satisfied by the Newton, quasi-Newton directions and the steepest descent direction with a nonsingular Jacobian.

**Theorem 3.2 [Global Convergence]**

Let Assumptions 3.1 be satisfied and let  $d_j$  satisfy  $\|W_j\| = \phi \|d_j\|$ , for some  $\phi \in (0, \infty)$ . Then the algorithms either terminate at a solution of the system  $W(y) = 0$  or they generate an infinite sequence  $\{y_j\}$  with a subsequence  $j \in J \subset \{0, 1, \dots\}$  such that  $\{\|d_j\|\} \rightarrow 0$  and thus every accumulation point  $y_*$  of the infinite sequence  $\{y_j\}$  is a solution of  $W(y) = 0$ .

**Proof**

By Lemmas (3.1) and (3.2), there exists a subsequence  $\{y_j\}$ ,  $j \in J$ , such that  $\|d_j\| \rightarrow 0$ . Let there exist  $y_*$  such that  $\{y_j\} \rightarrow y_*$ . The existence of such points is ensured since, the algorithms decrease  $\|W(y_j)\|_2^2$  at each iteration, thereby ensuring  $y_j \in \mathbb{F}$ , with  $\mathbb{F}$  compact. The result then follows by letting  $j \rightarrow \infty, j \in J$ , since

$$W(y_*) = \lim_{j \rightarrow \infty} W(y_j) = \lim_{j \rightarrow \infty} \phi \|d_j\| = 0. \quad \square$$

We demonstrate convergence to unit stepsizes in terms of a condition on the Jacobian or its approximation. To establish these results, we need to further assume that  $d_j$  is a Newton or quasi-(or approximate) Newton direction and strengthen the assumption on differentiability. We show that the convergence to unit stepsizes depends on the Jacobian. We first establish the result for an exact Jacobian and in Theorem 3.4 discuss the case for an approximate Jacobian such as a quasi-Newton approximation.

**Theorem 3.3 [Convergence to Unit Stepsizes - Exact Jacobian]**

Let Assumptions 3.1 be satisfied. Also let  $W \in C^2(\mathbb{R}^n)$  and

$$m \|v\|^2 \leq \langle v, [\nabla W_j]^T [\nabla W_j] v \rangle \leq M \|v\|^2; \quad m > 0; \quad \forall v \neq 0.$$

Then there is a stage  $J$  such that strategy (3.2) is satisfied with  $\tau_j = 1, \forall j \geq J$ .

**Proof**

Premultiplying the relationship  $\nabla W_j d_j = -W_j$  by  $W_j$ , we obtain

$$\langle ([\nabla W_j]^T W_j, d_j \rangle = - \langle W_j, W_j \rangle = - \langle d_j, [\nabla W_j]^T [\nabla W_j] d_j \rangle. \quad (3.6)$$

Consider the second order expansion of  $\|W\|_2^2$

$$\begin{aligned} \frac{1}{2} \| \mathcal{W}_{j+1} \|_2^2 - \frac{1}{2} \| \mathcal{W}_j \|_2^2 &= \langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle + \frac{1}{2} \langle \mathbf{d}_j, [\nabla \mathcal{W}_j]^T [\nabla \mathcal{W}_j] \mathbf{d}_j \rangle \\ &+ \int_0^1 (1-t) \langle \mathbf{d}_j, \left\{ \mathcal{Q}(y_j + t \mathbf{d}_j) - [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] \right. \\ &\left. + [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] - [\nabla \mathcal{W}_j]^T [\nabla \mathcal{W}_j] \right\} \mathbf{d}_j \rangle dt \quad (3.7) \end{aligned}$$

where  $\mathcal{Q}(\cdot)$  is the Hessian of  $\| \mathcal{W}(\cdot) \|_2^2$ , given by

$$[\nabla \mathcal{W}(\cdot)]^T [\nabla \mathcal{W}(\cdot)] + \sum_i \nabla^2 \mathcal{W}^i(\cdot) \mathcal{W}^i(\cdot),$$

with the last term vanishing at the solution  $\mathcal{W}(y_*) = 0$ . Using (3.6), (3.7) becomes

$$\frac{1}{2} \| \mathcal{W}_{j+1} \|_2^2 - \frac{1}{2} \| \mathcal{W}_j \|_2^2 \leq \langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle \left[ \frac{1}{2} - (\zeta_j + \frac{\xi_j}{2}) \right]$$

where

$$\zeta_j = \int_0^1 (1-t) \| \mathcal{Q}y(y_j + t \mathbf{d}_j) - [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] \| dt$$

$$\xi_j = \| [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] - [\nabla \mathcal{W}_j]^T [\nabla \mathcal{W}_j] \|.$$

Since, by Theorem 3.2,  $\{y_j\} \rightarrow y_*$ , we have  $\{\zeta_j\}, \{\xi_j\} \rightarrow 0$ . For  $\tau_j = 1$ , the scalar  $\mu \in [0, \frac{1}{2})$  of the stepsize strategy is bounded by

$$\mu \leq \frac{1}{2} - (\zeta_j + \frac{\xi_j}{2}).$$

Thus, when  $\frac{1}{2} - \mu \geq (\zeta_j + \frac{\xi_j}{2})$ , the stepsize strategy is satisfied with  $\tau_j = 1$ .  $\square$

The main difficulty about the use of approximate Jacobians is that  $\mathbf{d}_j$  does not always satisfy the descent condition  $\langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle \leq 0$ , and that  $\langle [\nabla \hat{\mathcal{W}}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle \leq 0$  does not necessarily imply descent. If descent is not ensured, then global convergence cannot be established. A better Jacobian approximation is required and the algorithms in the previous section aim to achieve this.

### **Theorem 3.4 [Convergence to Unit Stepsizes - Approximate Jacobian]**

Let Assumptions 3.1 be satisfied. Also let  $\mathcal{W} \in \mathbb{C}^2(\mathbb{R}^n)$  and

$$m \| \mathbf{v} \|^2 \leq \langle \mathbf{v}, [\nabla \hat{\mathcal{W}}_j]^T [\nabla \hat{\mathcal{W}}_j] \mathbf{v} \rangle \leq M \| \mathbf{v} \|^2; m > 0; \forall \mathbf{v} \neq 0,$$

and let  $\mathbf{d}_j$  that solves

$$\nabla \hat{\mathcal{W}}_j \mathbf{d} = - \mathcal{W}_j \quad (3.8)$$

satisfy the descent condition

$$\langle [\nabla \mathcal{W}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle \leq - \langle \mathcal{W}_j, \mathcal{W}_j \rangle. \quad (3.9)$$

Then the stepsize strategy is satisfied for some  $\tau_j \in (0, 1]$  and the monotonic decrease of the sequence  $\{ \|\mathcal{W}_j\|_2^2 \}$  is ensured. Also, there is a number  $\chi$  such that if

$$\boxed{\left\| [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] - [\nabla \hat{\mathcal{W}}_j]^T [\nabla \hat{\mathcal{W}}_j] \right\| \leq \chi} \quad (3.10)$$

then the stepsize strategy is satisfied for

$$\boxed{\tau_j = 1}.$$

**Proof**

Consider the second order expansion (3.7)

$$\begin{aligned} \frac{1}{2} \|\mathcal{W}_{j+1}\|_2^2 - \frac{1}{2} \|\mathcal{W}_j\|_2^2 &\leq \tau_j \langle [\nabla \hat{\mathcal{W}}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle \\ &+ \tau_j \langle [\nabla \mathcal{W}_j - \nabla \hat{\mathcal{W}}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle + \frac{1}{2} (\tau_j)^2 \langle \mathbf{d}_j, [\nabla \hat{\mathcal{W}}_j]^T [\nabla \hat{\mathcal{W}}_j] \mathbf{d}_j \rangle \\ &+ (\tau_j)^2 \left\{ \zeta_j + \frac{1}{2} \left\| [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] - [\nabla \hat{\mathcal{W}}_j]^T [\nabla \hat{\mathcal{W}}_j] \right\| \right\} \|\mathbf{d}_j\|^2 > dt \end{aligned} \quad (3.11)$$

From (3.8) and (3.9), we note that  $\langle [\nabla \mathcal{W}_j - \nabla \hat{\mathcal{W}}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle \leq 0$ . Also, from (3.8), we have

$$\langle ([\nabla \hat{\mathcal{W}}_j]^T \mathcal{W}_j, \mathbf{d}_j) \rangle = - \langle \mathcal{W}_j, \mathcal{W}_j \rangle = - \langle \mathbf{d}_j, [\nabla \hat{\mathcal{W}}_j]^T [\hat{\nabla} \mathcal{W}_j] \mathbf{d}_j \rangle \quad (3.12)$$

Using (3.12), (3.11) becomes

$$\begin{aligned} \frac{1}{2} \|\mathcal{W}_{j+1}\|_2^2 - \frac{1}{2} \|\mathcal{W}_j\|_2^2 &\leq \tau_j \langle [\nabla \hat{\mathcal{W}}_j]^T \mathcal{W}_j, \mathbf{d}_j \rangle \\ &\times \\ &\left[ 1 - \frac{1}{2} \tau_j \left[ 1 + \frac{1}{m} \left\{ 2 \zeta_j + \left\| [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] - [\nabla \hat{\mathcal{W}}_j]^T [\nabla \hat{\mathcal{W}}_j] \right\| \right\} \right] \right] \end{aligned}$$

Thus, there exists an  $\tau_j \in (0, 1]$  such that

$$\mu \leq 1 - \frac{1}{2} \tau_j \left[ 1 + \frac{1}{m} \left\{ 2 \zeta_j + \left\| [\nabla \mathcal{W}_*]^T [\nabla \mathcal{W}_*] - [\nabla \hat{\mathcal{W}}_j]^T [\nabla \hat{\mathcal{W}}_j] \right\| \right\} \right] \leq \frac{1}{2} \quad (3.13)$$

which satisfies the stepsize strategy. The monotonic decrease of the sequence  $\{ \|\mathcal{W}_j\|_2^2 \}$  follows from this property. Global convergence follows from Theorem 3.2. Thus,  $\{y_j\} \rightarrow y_*$  and  $\{\zeta_j\} \rightarrow 0$ .

We can use (3.13) to conclude that if  $\chi > 0$  is such that

$$\frac{1}{2} - \mu \geq \left[ \frac{1}{m} \left\{ 2 \zeta_j + \chi \right\} \right]$$

(in view of  $\{\zeta_j\} \rightarrow 0$ , this defines the number  $\chi$ ) then the stepsize strategy holds with  $\tau_j = \square$

**Lemma 3.3**

Let  $\{y^j\} \rightarrow y^*$ , then  $\{y^j\}$  is Q-superlinearly convergent if and only if

$$\boxed{\|d^j\| \leq \gamma^j \|d^{j-1}\|, \lim_{j \rightarrow \infty} \gamma^j = 0.}$$

**Proof**

Suppose  $\|d_j\| \leq \gamma_j \|d_{j-1}\|$ ,  $\lim_{j \rightarrow \infty} \gamma_j = 0$ , holds. We have

$$\begin{aligned} \|y_* - y_j\| &\leq \lim_{t \rightarrow \infty} \sum_{i=j}^{t-1} \|y_{i+1} - y_i\| \leq \gamma_j \|d_{j-1}\| (1 + \omega + \omega^2 + \omega^3 + \dots) \\ &\leq \frac{\gamma_j}{1-\omega} \left\{ \|y_j - y_*\| + \|y_* - y_{j-1}\| \right\} \end{aligned}$$

for some  $\omega \in [0, 1)$ . As  $\{\gamma_j\} \rightarrow 0$ ,  $\omega$  is chosen such that  $\gamma_j + \omega < 1, \forall j \geq J_0$ .  $J_0$  is an integer and is such that  $\gamma_j < 1, \forall j \geq J_0$ . Rearranging the above expression,

$$\begin{aligned} \frac{\|y_j - y_*\|}{\|y_j - y_*\| + \|y_{j-1} - y_*\|} &\leq \frac{\gamma_j}{1-\omega} \\ \|y_* - y_j\| &\leq \frac{\gamma_j}{1-\omega-\gamma_j} \|y_* - y_{j-1}\| \end{aligned} \tag{3.14}$$

which establishes the Q-superlinear convergence of  $\{y_j\}$ .

Suppose that  $\|y_* - y_j\| \leq \beta_j \|y_* - y_{j-1}\|$ ,  $\lim_{j \rightarrow \infty} \beta_j = 0$ , with  $\beta_j < 1$ . This yields the inequality

$$\begin{aligned} \|y_* - y_j\| &\leq \beta_j \|y_* - y_j\| + \beta_j \|d_{j-1}\| \\ &\leq \left( \frac{\beta_j}{1-\beta_j} \right) \|d_{j-1}\| \end{aligned} \tag{3.15}$$

Next, consider

$$-\nabla \hat{\mathcal{W}}_j d_j = \mathcal{W}_j = \mathcal{W}_* + \int_0^1 \nabla \mathcal{W}(y_j + t(y_* - y_j)) (y_* - y_j) dt$$

which yields, for  $c_1 \in [0, \infty)$ ,

$$\|d_j\| \leq c_1 \|y_* - y_j\|. \tag{3.16}$$

Using (3.16) in (3.15), leads to the required result

$$\|d_j\| \leq c_1 \left( \frac{\beta_j}{1-\beta_j} \right) \|d_{j-1}\|. \quad \square$$

**Theorem 3.5**

Let  $D \subset \mathbb{R}^n$  be an open convex set, Assumptions 3.1 be satisfied and let  $\{\nabla \hat{\mathcal{W}}_j\}$  be a sequence of nonsingular matrices and suppose that for some  $y_0 \in D$  the sequence generated by

$$\boxed{y_{j+1} = y_j + d_j \quad \text{with} \quad \nabla \hat{\mathcal{W}}_j d_j = -\mathcal{W}_j} \quad (3.17)$$

remains in  $D$ , and  $\lim_{j \rightarrow \infty} y_j = y_*$ . Then  $\{y_j\}$  satisfies  $\|d_j\| \leq \gamma_j \|d_{j-1}\|$ ,  $\lim_{j \rightarrow \infty} \gamma_j = 0$ , and thence converges Q-superlinearly to  $y_*$ , in some norm  $\|\cdot\|$  and  $\mathcal{W}(y_*) = 0$  iff

$$\boxed{\lim_{j \rightarrow \infty} \frac{\|[\nabla \hat{\mathcal{W}}_j - \nabla \mathcal{W}_*] d_j\|}{\|d_j\|} = 0}$$

### Proof

In order to establish this result in both directions, we need only to consider (3.17) and the first order expansion of  $\mathcal{W}(y)$

$$-\nabla \hat{\mathcal{W}}_j d_j = \mathcal{W}_j = \mathcal{W}_{j-1} + \nabla \hat{\mathcal{W}}_{j-1} d_{j-1} + \Psi_j + [\nabla \mathcal{W}_* - \nabla \hat{\mathcal{W}}_{j-1}] d_{j-1}$$

where

$$\Psi_j = \int_0^1 [\nabla \mathcal{W}(y(t)) - \nabla \mathcal{W}_*] d_{j-1} dt,$$

with  $\{\Psi_j\} \rightarrow 0$  and  $\mathcal{W}_{j-1} + \nabla \hat{\mathcal{W}}_{j-1} d_{j-1} = 0$ . As  $\nabla \hat{\mathcal{W}}_j$  is nonsingular, dividing the above expression by  $\|d_{j-1}\|$  yields the required result.  $\square$

### REFERENCES

- Başar, T. (1987). "Relaxation Techniques and Asynchronous Algorithms for on-line Computation of Non-cooperative Equilibria", *Journal of Economic Dynamics and Control*, 11, 531-550.
- Başar, T and Olsder, G.L. (1982). *Dynamic Noncooperative Games*, Academic Press, New York.
- Becker, R. and B. Rustem (1993). "Algorithms for Solving Nonlinear Dynamic Decision Models", *Annals of Operations Research*, 44, 117-142.
- Bertsekas, D.P. (1979). "Convexification Procedures and Decomposition methods for Non-convex Optimisation Problems", *JOTA*, 29, 169-197.
- Bertsekas, D.P. and P. Tseng (1994). "Partial Proximal Minimization Algorithms for Convex Programming", *SIAM J Optimization*, 4, 551-572.
- Bertsekas, D.P. and J.N. Tsitsiklis (1989). *Parallel and Distributed Computing*, Prentice-Hall International, New Jersey.
- Dennis Jr., J.E. and J.J. Moré (1974). "A Characterisation of Superlinear Convergence and its Application to Quasi-Newton Methods", *Maths. Comp.*, 28, 549-560.

- Dennis Jr., J.E. and R.B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall Inc, New Jersey.
- Ho, Y.C. (1970). "Differential Games, Dynamic Optimization and Generalized Control Theory", *JOTA*, 6, 179-209.
- Khachiyan, L.G. and A.I. Erlich (1978). "Serial Game Processes of Solution of Convex Programming Problems", *Engineering Cybernetics*, 16, 147-151
- Li, G. (1989). "Successive Column Correction Algorithms for Solving Sparse Nonlinear Systems of Equations", *Mathematical Programming*, 43, 187-208.
- More', J.J. and J.A. Trangenstein (1976). "On the Global Convergence of Broyden's Method", *Mathematics of Computation*, 30, 523-540.
- Ortega, J. (1972). *Numerical Analysis: A Second Course*, Academic Press, New York and London.
- Ortega, J. and W.C. Rheinboldt (1970). *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York and London.
- Rockafeller, R.T. (1976). "Monotone Operators and the Proximal Point Algorithm", *SIAM J Control Optim*, 14, 877-898.
- Rosen, J.B. (1965). "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games", *Econometrica*, 33, 520-534.