DAMTP 2011/NA01

On the convergence of trust region algorithms for unconstrained minimization without derivatives[1]

M.J.D. Powell

Abstract: We consider iterative trust region algorithms for the unconstrained minimization of an objective function $F(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, when F is differentiable but no derivatives are available, and when each model of F is a linear or a quadratic polynomial. The models interpolate F at $n+1$ points, which defines them uniquely when they are linear polynomials. In the quadratic case, second derivatives of the models are derived from information from previous iterations, but there are so few data that typically only the magnitudes of second derivative estimates are correct. Nevertheless, numerical results show that much faster convergence is achieved when quadratic models are employed instead of linear ones. Just one new value of F is calculated on each iteration. Changes to the variables are either trust region steps or are designed to maintain suitable volumes and diameters of the convex hulls of the interpolation points. It is proved that, if F is bounded below, if $\nabla^2 F$ is also bounded, and if the number of iterations is infinite, then the sequence of gradients $\underline{\nabla} F(\underline{x}_k)$, $k = 1, 2, 3, \ldots$, converges to zero, where $\underline{x}_k$ is the centre of the trust region of the k-th iteration.

Keywords: Convergence theory; Derivative free optimization; Symmetric Broyden; Trust region methods; Unconstrained minimization.

Department of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge CB3 0WA,
England.

January, 2011.

[1]Presented at the Workshop on Nonlinear Optimization, Variational Inequalities and Equilibrium Problems (Erice, Italy, 2010).

1. Introduction

We study the convergence of some iterative algorithms for seeking the least value of a function $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, which is defined by a subroutine that returns the value $F(\underline{x})$ for any given vector of variables $\underline{x}\in\mathcal{R}^n$. The convergence analysis requires F to be bounded below and to have bounded second derivatives, but no derivatives are available to the algorithms. At the beginning of every iteration of an algorithm, there is a quadratic (or linear) polynomial function

$$Q_k(\underline{x}) \;=\; F(\underline{x}_k)+(\underline{x}-\underline{x}_k)^T\underline{g}_k+\tfrac{1}{2}\,(\underline{x}-\underline{x}_k)^TG_k\,(\underline{x}-\underline{x}_k),\qquad \underline{x}\in\mathcal{R}^n, \tag{1.1}$$

that is employed as an approximation to $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, when $\underline{x}$ is sufficiently close to $\underline{x}_k$, where k is the iteration number, and where $\underline{x}_k$ is the vector of variables that has supplied the least calculated value of the objective function so far. The vector $\underline{g}_k\in\mathcal{R}^n$ and the $n\times n$ symmetric matrix G_k, which may be zero, are generated by the algorithms. The k-th iteration picks a nonzero step $\underline{d}_k$ from $\underline{x}_k$, and then calls the subroutine that returns the new function value $F(\underline{x}_k+\underline{d}_k)$. There are two kinds of step $\underline{d}_k$, namely "trust region" steps, chosen to make $Q_k(\underline{x}_k+\underline{d}_k)$ substantially less than $Q_k(\underline{x}_k)$, in the hope that most of this reduction in Q_k may be inherited by F, and "alternative" steps, designed to make the approximations $Q_k\approx F$, $k=1,2,3,\ldots$, sufficiently accurate. The k-th iteration also picks the new approximation Q_{k+1}, which satisfies $Q_{k+1}(\underline{x}_k+\underline{d}_k)=F(\underline{x}_k+\underline{d}_k)$, and it applies the formula

$$\underline{x}_{k+1} \;=\; \begin{cases} \underline{x}_k, & F(\underline{x}_k+\underline{d}_k)\geq F(\underline{x}_k)\\ \underline{x}_k+\underline{d}_k, & F(\underline{x}_k+\underline{d}_k)<F(\underline{x}_k).\end{cases} \tag{1.2}$$

There are at least three important differences between descriptions of algorithms for practical use and descriptions for convergence theory. One is that in practice the finite precision of computer arithmetic requires careful attention, but we make the usual assumptions that all computations are exact, and that the number of iterations can be infinite. Secondly, all questions about the details of a practical algorithm have to be answered specifically, but generality is welcome in convergence theory, in order to broaden the range of applicability of the analysis. Thirdly, algorithms for practical use may include techniques that are successful in numerical experiments, but that are without conditions that are needed at present for proofs of convergence. Thus the NEWUOA software (Powell, 2006), for example, is much more efficient in practice than the methods studied below.

Throughout this paper, every approximation Q_k satisfies $n+1$ interpolation conditions

$$Q_k(\underline{y}_i) \;=\; F(\underline{y}_i),\qquad i=0,1,\ldots,n, \tag{1.3}$$

where $\underline{y}_0$ is the point $\underline{x}_k$, and where all the values $F(\underline{y}_i)$, $i=0,1,\ldots,n$, have been computed already, either in some preliminary work or on previous iterations. In the NEWUOA software, however, more than $n+1$ interpolation conditions are employed, in order to supply some information about second derivatives of F. It is crucial to our theory that, after choosing the matrix G_k of expression (1.1),

the equations (1.3) provide the gradient $\underline{\nabla} Q_k(\underline{x}_k) = \underline{g}_k$ uniquely. An equivalent statement of this nondegeneracy condition is that the $(n+1)\times(n+1)$ matrix

$$D_k \;=\; \left(\begin{array}{cccc} \underline{y}_0 & \underline{y}_1 & \cdots & \underline{y}_n \\ 1 & 1 & \cdots & 1 \end{array} \right) \tag{1.4}$$

is nonsingular. Another equivalent statement is that the volume of the convex hull in $\mathcal{R}^n$ of the points $\underline{y}_i$, $i=0,1,\ldots,n$, is nonzero, the volume of the convex hull being $|\det D_k| / n!$.

In our analysis, every G_k is a bounded symmetric matrix, and, except in one reasonable situation that is addressed later, every G_k can be chosen arbitrarily. Then the vector $\underline{g}_k$ of expression (1.1) is defined by the equations (1.3) with $Q_k(\underline{x}_k) = F(\underline{x}_k) = F(\underline{y}_0)$. We allow G_k to be nonzero, because often the progress of iterative algorithms for unconstrained optimization is unacceptably slow if the curvature of the objective function is ignored. A convenient way of choosing a nonzero matrix G_{k+1} automatically is described at the end of this paper. It obtains some second derivative information by combining the new calculated function value $F(\underline{x}_k + \underline{d}_k)$ with the available function values $F(\underline{y}_i)$, $i=0,1,\ldots,n$. The reader may find it helpful initially, however, to take the view that all the matrices G_k, $k=1,2,3,\ldots$, are zero.

The convergence of algorithms for optimization without derivatives receives much attention in the works of Conn, Scheinberg and Vicente (1997, 2009a, 2009b). They have developed most of the published theory of derivative-free methods that take trust region steps, using a linear or quadratic approximation to F on each iteration. Let $\rho_k > 0$ be the trust region radius of the k-th iteration, which means that, when the step $\underline{d}_k$ of the new function value $F(\underline{x}_k + \underline{d}_k)$ is picked, it has to satisfy $\|\underline{d}_k\| \le \rho_k$. They address linear approximations $Q_k \approx F$, defined by interpolation equations of the form (1.3), and they explain the importance of the conditions

$$\|\underline{y}_i - \underline{x}_k\| \;\le\; c_1 \rho_k, \qquad i=1,2,\ldots,n, \tag{1.5}$$

and

$$|\det D_k| \;\ge\; c_2 \rho_k^n, \tag{1.6}$$

where c_1 and c_2 are positive constants and where D_k is the matrix (1.4). These conditions are also employed by the COBYLA algorithm of Powell (1994).

Inequalities (1.5) and (1.6), with the boundedness of $\nabla^2 F$ and $\nabla^2 Q_k$, imply that the gradient $\underline{g}_k = \underline{\nabla} Q_k(\underline{x}_k)$ of the function (1.1) has the property

$$\|\, \underline{\nabla} Q_k(\underline{x}_k) - \underline{\nabla} F(\underline{x}_k) \,\| \;\le\; c_3 \rho_k, \tag{1.7}$$

where c_3 is another positive constant. A proof of this assertion is given in Conn *et al* (2009b) and at the end of Section 3 below. Further, if $\|\underline{\nabla} F(\underline{x}_k)\|$ is much larger than $c_3 \rho_k$, and if $\underline{d}_k$ is a "trust region" step, then condition (1.7) implies that the relative error of the approximation $F(\underline{x}_k) - F(\underline{x}_k + \underline{d}_k) \approx Q_k(\underline{x}_k) - Q_k(\underline{x}_k + \underline{d}_k)$ is tiny. In other words, most of the reduction in Q_k due to the "trust region"

step is inherited by F. In this situation, standard methods for adjusting the trust region radius provide $\rho_{k+1} \geq \rho_k$. It follows that, when $\rho_{k+1} < \rho_k$ occurs, then $\|\underline{\nabla} F(\underline{x}_k)\|$ is also bounded above by a constant multiple of ρ_k. We define the set $\mathcal{K}$ by including the iteration number k in $\mathcal{K}$ if and only if ρ_{k+1} is less than all the numbers ρ_j, $j=1,2,\ldots,k$. Hence, if the number of elements of $\mathcal{K}$ is infinite, and if the decreasing sequence ρ_{k+1}, $k\in\mathcal{K}$, tends to zero, then the sequence $\|\underline{\nabla} F(\underline{x}_k)\|$, $k \in \mathcal{K}$, tends to zero too. This argument provides the gist of our convergence theory. Furthermore, it is proved in Section 7 that our algorithms supply the limit

$$\| \underline{\nabla} F(\underline{x}_k) \| \to 0 \quad \text{as} \quad k \to \infty, \tag{1.8}$$

when k runs through all the positive integers. An unusual feature of this result is that the choice (1.2) of $\underline{x}_{k+1}$ is without a "sufficient decrease" condition.

There is a major strategic difference between the method of COBYLA (Powell, 1994) and the methods of Conn *et al* (1997, 2009a, 2009b) for achieving the bounds (1.5) and (1.6). In COBYLA, and in the algorithms of our convergence theory, just one new function value $F(\underline{x}_k+\underline{d}_k)$ is calculated on each iteration, where $\underline{d}_k$ is either a "trust region" or an "alternative" step, as mentioned in the opening paragraph of this section. Then the set $\{\underline{y}_i : i=0,1,\ldots,n\}$ of interpolation points for the next iteration is formed by picking an integer t from $[1,n]$, and by replacing the old $\underline{y}_t$ by $\underline{x}_k+\underline{d}_k$, all the other old interpolation points being retained. Thus only the $(t+1)$-th column of the matrix (1.4) is altered, except that the first and $(t+1)$-th columns of the new D_k are exchanged if $F(\underline{x}_k+\underline{d}_k)<F(\underline{x}_k)$ occurs, which preserves the property that $F(\underline{y}_0)$ is the least calculated function value so far. It is proved in Sections 4 and 5 that our "alternative" steps give the conditions (1.5) and (1.6), these steps being designed either to move a point $\underline{y}_t$ that is unacceptably far from $\underline{x}_k$ or to provide a substantial increase in $|\det D_k|$. We apply the strategy that, if $\underline{d}_k$ is an "alternative" step, then usually $\underline{d}_{k+1}$ is a "trust region" step, our aim being to take advantage immediately of any improvement to the approximation $Q_k \approx F$. On the other hand, Conn *et al* maintain the bounds (1.5) and (1.6) by employing a "model-improvement" algorithm, which calculates additional values of F if necessary. Whenever it is invoked, it guarantees that all of the conditions (1.5) and (1.6) are satisfied, so it can happen that many new values of the objective function are computed without a "trust region" step. Our approach may be much more efficient when only a small change in $Q_k \approx F$ is sufficient for the success of the next "trust region" step.

Our set of algorithms is specified in Section 2. Attention is given to the matrix (1.4) in Section 3, with the analysis that provides inequality (1.7). The achievement of conditions (1.5) and (1.6) by our algorithms is established in Sections 4 and 5, respectively. We find in Section 6 that, when the number of iterations is infinite, at least a subsequence of the norms $\|\underline{\nabla} F(\underline{x}_k)\|$, $k=1,2,3,\ldots$, tends to zero. The limit (1.8) is proved in Section 7. Numerical experiments that investigate some nonzero choices of the matrices $G_k=\nabla^2 Q_k$, $k=1,2,3,\ldots$, are reported and discussed in Section 8.

2. Specification of the algorithms

The interpolation points $\underline{y}_i$, $i=0,1,\ldots,n$, and the second derivative matrix $G_1=\nabla^2 Q_1$ of the first quadratic model $Q_1\approx F$ are chosen before the first iteration. The only restriction on the initial points is that the volume of their convex hull in $\mathcal{R}^n$ has to be positive, which means that the initial matrix (1.4) is nonsingular. The only restriction on G_1 is that the matrices G_k, $k=1,2,3,\ldots$, have to be symmetric and uniformly bounded. After calculating the function values $F(\underline{y}_i)$, $i=0,1,\ldots,n$, the points are reordered if necessary to satisfy the conditions

$$F(\underline{y}_0)\ \leq\ F(\underline{y}_i),\qquad i=1,2,\ldots,n. \tag{2.1}$$

We recall that every model has the form (1.1), where $\underline{x}_k=\underline{y}_0$, and where the gradient $\underline{g}_k\in\mathcal{R}^n$ is defined by the equations (1.3) after G_k has been fixed.

We let the trust region radii take values that simplify the theory. The initial radius ρ_1 can be any positive number. We set $\rho=\rho_1$, where $\rho=\rho_1$ is a lower bound on ρ_k during the early iterations, the condition $\rho_k\geq\rho$ being required until it seems that a reduction in ρ is necessary for further progress, as described later. It is possible to prove the given lemmas and theorems when each ρ_k is any number from the interval $[\rho,M\rho]$, where M is a constant that satisfies $M\geq 1$, but we make the simplification $M=1$, although $\rho_k>\rho$ is needed for efficiency in practice if ρ_1 is unsuitably small. Specifically, the current trust region radius ρ is never increased, but it is reduced occasionally, the condition $\rho\to 0$ as $k\to\infty$ being important to the proof of convergence. Another simplification is that every reduction in ρ is by a factor of 10, but instead each factor could be any number from the interval $[M_1,M_2]$, where M_1 and M_2 are any constants such that $1<M_1\leq M_2$. For each iteration number k, we let κ be the number of the first iteration that is provided with the current value of ρ.

The "Cauchy" step, $\underline{\widehat{d}}_k$ say, of the model (1.1) is defined to be the multiple of the gradient $\underline{g}_k$ that minimizes $Q_k(\underline{x}_k+\underline{\widehat{d}}_k)$ subject to $\|\underline{\widehat{d}}_k\|\leq\rho$, with $\underline{\widehat{d}}_k=0$ if $\underline{g}_k$ is zero. Every "trust region" step of our algorithms is allowed to be any vector $\underline{d}_k$ that has the properties

$$Q_k(\underline{x}_k+\underline{d}_k)\ \leq\ Q_k(\underline{x}_k+\underline{\widehat{d}}_k)\qquad\text{and}\qquad\|\underline{d}_k\|\ \leq\ \rho, \tag{2.2}$$

except that the step must be "exact" if $k\geq\kappa+5$ and if the number

$$\eta_k\ =\ \max\{\,|Q_j(\underline{x}_j+\underline{d}_j)-F(\underline{x}_j+\underline{d}_j)|:j=\kappa,\kappa+1,\ldots,k-1\,\} \tag{2.3}$$

is zero, the meaning of "exact" being that $\underline{d}_k$ has to be the vector $\underline{d}$ that actually minimizes $Q_k(\underline{x}_k+\underline{d})$ subject to $\|\underline{d}\|\leq\rho$. We see that η_k is the greatest error of the approximation $Q_j(\underline{x}_j+\underline{d}_j)\approx F(\underline{x}_j+\underline{d}_j)$, as j runs through the numbers of the iterations that have been completed already with the current value of ρ. We define η_κ to be zero for use later. Our theory remains valid if the condition $k\geq\kappa+5$ for an "exact" trust region step is replaced by $k\geq\kappa+M_3$, where M_3

is any constant positive integer, but we prefer to be parsimonious in our use of parameters. Another case of parsimony occurs in the inequality

$$F(\underline{x}_k) - F(\underline{x}_k + \underline{d}_k) \;\geq\; 0.1\,\{Q_k(\underline{x}_k) - Q_k(\underline{x}_k + \underline{d}_k)\}, \tag{2.4}$$

the factor 0.1 being replaceable by any positive constant that is less than one. We say that a "trust region" step is "successful" if and only if it achieves a reduction in the objective function that is bounded below by condition (2.4).

The choice of every "alternative" step $\underline{d}_k$ is made after the integer t has been picked from $[1,n]$, where $\underline{y}_t$ is still the old interpolation point that is going to be dropped to make room for $\underline{x}_k+\underline{d}_k$. Then $\underline{d}_k$ is defined, except for its sign, by maximizing the volume of the convex hull of the new set of interpolation points in $\mathcal{R}^n$, subject to $\|\underline{d}_k\| \leq \rho$. It follows that the direction of $\underline{d}_k$ from $\underline{x}_k = \underline{y}_0$ is orthogonal to the face of the convex hull that has the vertices $\underline{y}_i$, $i \in \{0,1,\ldots,n\}\backslash\{t\}$, and the length of $\underline{d}_k$ is ρ. The sign of this step does not matter, but it is suitable to prefer the sign that provides the smaller value of $Q_k(\underline{x}_k+\underline{d}_k)$. There are two kinds of "alternative" steps, namely "alpha" and "beta" steps, that differ in their choice of t; they supply the conditions (1.6) and (1.5), respectively.

Our algorithms require five real parameters to be set in advance, namely α, β, γ, τ_α and τ_β. The value of α can be any constant from the open interval $(0,1)$, while β and γ can be any numbers that satisfy $\beta>1$ and $\gamma>0$. Both τ_α and τ_β can be any positive integers. The settings $\alpha=0.1$, $\beta=5$, $\gamma=0.01$, $\tau_\alpha=1$ and $\tau_\beta=5$ are going to be used in the numerical experiments of Section 8. The decision whether or not to take an "alpha" or a "beta" step depends on α and τ_α or on β and τ_β, respectively, while γ is employed in the decision whether or not to take a "trust region" step, as explained below.

The index t of an "alpha" step is given the value that maximizes the volume of the convex hull of the new set of interpolation points. For $i = 1,2,\ldots,n$, let σ_i be the distance from $\underline{y}_i$ to the hyperplane that contains the points $\underline{y}_j$, $j \in \{0,1,\ldots,n\}\backslash\{i\}$, all of the points being the ones at the beginning of the current iteration. The replacement of $\underline{y}_t$ by $\underline{x}_k+\underline{d}_k$, where $\underline{d}_k$ is an "alternative" step, multiplies the volume of the convex hull by ρ/σ_t. Therefore, if the taking of an "alpha" step is under consideration, the numbers σ_i, $i = 1,2,\ldots,n$, are calculated, and t is set to any integer from $[1,n]$ that satisfies $\sigma_t \leq \sigma_i$, $i=1,2,\ldots,n$, which usually defines t uniquely. There is no need for an "alpha" step, however, if σ_t is sufficiently large. Therefore the step is actually taken if and only if the condition

$$\sigma_t \;=\; \min\{\sigma_i : i=1,2,\ldots,n\} \;<\; \alpha\,\rho \tag{2.5}$$

holds, which shows the purpose of the parameter $\alpha \in (0,1)$. The purpose of τ_α is that, if τ_α "trust region" steps have been taken with the current ρ since the last attempt at an "alpha" step, then the procedure of this paragraph must be applied before the next attempt at a "trust region" step.

A "beta" step replaces the interpolation point $\underline{y}_t$ by $\underline{x}_k+\underline{d}_k$ because $\|\underline{y}_t-\underline{x}_k\|$ is substantially larger than ρ, but the choice of t from $[1,n]$ is not always the integer

i that gives the greatest of the distances $\|\underline{y}_i-\underline{x}_k\|$, $i=1,2,\ldots,n$. By omitting some of these values of i, the conditions for taking a "beta" step become helpful to our criterion for terminating the iterations with the current value of ρ. Thus termination is guaranteed if a sufficiently long sequence of consecutive iterations is without a "trust region" step that achieves the "success" of inequality (2.4). The technique employs a set $\mathcal{B}$ of the integers $\{1,2,\ldots,n\}$, all of these integers being included in $\mathcal{B}$ at the beginning of each iteration with a new value of ρ and whenever a "trust region" step is "successful", but otherwise the number of elements of $\mathcal{B}$ decreases monotonically. With every replacement of $\underline{y}_t$ by $\underline{x}_k+\underline{d}_k$, the integer t is deleted from $\mathcal{B}$ unless it has been removed already. When a "beta" step is under consideration, and when $\mathcal{B}$ is not empty, the integer t is set to an element of $\mathcal{B}$ that has the property $\|\underline{y}_t-\underline{x}_k\|\geq\|\underline{y}_i-\underline{x}_k\|$, $i\in\mathcal{B}$. Then $\underline{y}_t$ is replaced by $\underline{x}_k+\underline{d}_k$, where $\underline{d}_k$ is an "alternative" step, if and only if $\mathcal{B}$ is not empty and the inequality

$$\|\underline{y}_t-\underline{x}_k\| \;=\; \max\{\|\underline{y}_i-\underline{x}_k\|:i\in\mathcal{B}\} \;>\; \beta\rho \tag{2.6}$$

holds, which shows the purpose of the parameter $\beta>1$. The purpose of τ_β is that, if τ_β "trust region" steps have been taken with the current ρ since the last attempt at a "beta" step, then the procedure of this paragraph must be applied before the next attempt at a "trust region" step.

It is possible not only for "alpha" and "beta" steps to be considered and not taken, but also for "trust region" steps to be generated and then abandoned. We employ the number (2.3), with $\eta_\kappa=0$, to estimate the accuracy of the approximation $Q_k(\underline{x}_k+\underline{d}_k)\approx F(\underline{x}_k+\underline{d}_k)$, and we assume that a "trust region" step $\underline{d}_k$ is likely to be useful only if the predicted reduction $Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)$ compares favourably with η_k. Moreover, the length of a step may suggest that the time has come for a decrease in ρ. Therefore a "trust region" step $\underline{d}_k$ is taken, the new function value $F(\underline{x}_k+\underline{d}_k)$ being calculated, if and only if the conditions

$$Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k) \;>\; \gamma\,\eta_k \qquad\text{and}\qquad \|\underline{d}_k\| \;\geq\; \tfrac{1}{2}\,\rho, \tag{2.7}$$

are satisfied, which shows the purpose of the parameter $\gamma>0$. The property $\eta_\kappa=0$ is a disadvantage, but our theory would become more difficult if the definition (2.3) of η_k were augmented by errors $|Q_j(\underline{x}_j+\underline{d}_j)-F(\underline{x}_j+\underline{d}_j)|$ with j less than κ.

The choice of the old interpolation point $\underline{y}_t$ that is dropped to make room for $\underline{x}_k+\underline{d}_k$, after calculating the new function value $F(\underline{x}_k+\underline{d}_k)$, is given above for "alternative" steps $\underline{d}_k$. The following technique is applied when $\underline{d}_k$ is a "trust region" step. Our theory requires the volume of the new convex hull of interpolation points to be bounded below by the volume of the old convex hull multiplied by a positive constant, which is established later for our choice of t. Specifically, after $F(\underline{x}_k+\underline{d}_k)$ is calculated for a "trust region" step $\underline{d}_k$, we let the multipliers θ_i, $i=0,1,\ldots,n$, satisfy the equations

$$\underline{x}_k+\underline{d}_k \;=\; \sum_{i=0}^{n}\theta_i\,\underline{y}_i \qquad\text{and}\qquad \sum_{i=0}^{n}\theta_i \;=\; 1, \tag{2.8}$$

which defines them uniquely, due to the nonsingularity of the matrix (1.4). Then t is set to any integer from $[1,n]$ that has the property

$$|\theta_t| \geq |\theta_i|, \qquad i=1,2,\ldots,n. \tag{2.9}$$

Not all of the numbers θ_i, $i=1,2,\ldots,n$, are zero, because, if they were, then the equations (2.8) would imply $\underline{x}_k+\underline{d}_k=\underline{y}_0$, which would give the contradiction $\underline{d}_k=0$. We find in Section 3 that the volume of the new convex hull is the volume of the old convex hull multiplied by $|\theta_t|$.

Nearly all of the conditions on the new model $Q_{k+1}(\underline{x})\approx F(\underline{x})$, $\underline{x}\in\mathcal{R}^2$, have been stated already. Because of the interpolation equations, all the freedom in Q_{k+1} is given by the freedom in the new second derivative matrix $G_{k+1}=\nabla^2 Q_{k+1}$. We recall from Section 1, however, that there is one "reasonable situation" when G_{k+1} cannot be chosen arbitrarily, subject to symmetry and uniform boundedness. This situation is related intimately to the need for "exact" trust region steps, introduced in the sentence that includes expressions (2.2) and (2.3), and explained in Section 6. It occurs if both $k\geq\kappa+5$ and $\eta_{k+1}=0$ hold, and then it is mandatory to pick $G_{k+1}=G_k$, except that, as stated already, $k\geq\kappa+5$ can be replaced by $k\geq\kappa+M_3$, where M_3 is any constant positive integer. The value $\eta_{k+1}=0$ includes $Q_k(\underline{x}_k+\underline{d}_k)=F(\underline{x}_k+\underline{d}_k)$, so in this case $G_{k+1}=G_k$ implies $Q_{k+1}\equiv Q_k$. In other words, we retain the old model when there is no need for a change.

We recall that just one new function value $F(\underline{x}_k+\underline{d}_k)$ is computed on each iteration, but that $F(\underline{x}_k+\underline{d}_k)$ is not calculated if condition (2.5), (2.6) or at least one of the inequalities (2.7) fails when trying to take an "alpha", "beta" or "trust region" step, respectively. Thus more than one kind of step may be tried on an iteration. We prefer "trust region" steps to occur frequently. Therefore we impose the rule that not more than one attempt at an "alpha" step and not more than one attempt at a "beta" step are allowed between any two consecutive attempts at a "trust region" step with the same value of ρ. Furthermore, we say that a "trust region" attempt is "unsuccessful" if $F(\underline{x}_k+\underline{d}_k)$ is not calculated or if the reduction (2.4) is not achieved, and then we require a "beta" step to be tried before the next attempt at a "trust region" step. Moreover, the first and second attempts at steps with each new value of ρ are of "alpha" type and of "trust region" type, respectively. There are no more restrictions on the steps of the algorithms for our convergence theory, although some choices are still open. For example, the algorithm in Section 8 attempts an "alpha" step after every attempt at a "trust region" step, but this feature is unnecessary in our analysis of convergence.

The iterations with the current value of ρ are terminated if a "trust region" step is "unsuccessful", as defined in the previous paragraph, and if the next attempt at a "beta" step does not alter the interpolation points, either because $\mathcal{B}$ is empty or because all of the distances $\|\underline{y}_i-\underline{x}_k\|$, $i\in\mathcal{B}$, are at most $\beta\rho$. It is proved in Section 6 that termination occurs for every ρ in exact arithmetic, even if the number (2.3) does not become positive as k increases. We find also that, when the calculations with the current ρ are complete, then $\|\underline{\nabla}F(\underline{x}_k)\|$ is bounded above by a constant multiple of ρ, as stated in Section 1.

3. Some properties of the interpolation matrix (1.4)

Inequality (1.7) is established at the end of this section, under the conditions stated in Section 1, which include the bounds (1.5) and (1.6), where c_1, c_2 and c_3 are positive constants. As in Section 2, we drop the subscript k from the notation ρ_k for the trust region radius, which reminds us that ρ may not be altered during many consecutive iterations. Our theory employs the Lagrange functions Λ_j, $j=0,1,\ldots,n$, of the interpolation equations (1.3), where $\Lambda_j(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is the linear polynomial that takes the values

$$\Lambda_j(\underline{y}_i) \;=\; \delta_{ij}, \qquad i=0,1,\ldots,n, \tag{3.1}$$

the right hand side δ_{ij} being the Kronecker delta. The coefficients of each Λ_j are defined uniquely by the equations (3.1), because every matrix (1.4) is nonsingular. We require the remark that the conditions (1.5) and (1.6) imply the property

$$\textstyle\sum_{j=0}^n\,|\Lambda_j(\underline{x})| \;\leq\; c_4 \qquad \text{if} \qquad \|\underline{x}-\underline{x}_k\| \;\leq\; \rho, \tag{3.2}$$

where c_4 is another positive constant. It is going to be derived from the dependence of the Lagrange functions on the determinant of the matrix (1.4), after establishing the upper bound

$$|\det D_k| \;\leq\; \textstyle\prod_{i=1}^n\,\|\underline{y}_i-\underline{y}_0\| \tag{3.3}$$

and the lower bound

$$|\det D_k| \;\geq\; \textstyle\prod_{i=1}^n\,\sigma_i, \tag{3.4}$$

where we recall from Section 2 that σ_i is the distance from $\underline{y}_i$ to the hyperplane that contains the points $\underline{y}_j$, $j\in\{0,1,\ldots,n\}\backslash\{i\}$. This analysis suggests a convenient way of generating the "alternative" steps $\underline{d}_k$, specified in the first whole paragraph after expression (2.4).

The justification of the bounds (3.3) and (3.4) begins with the identity

$$\begin{aligned}\det D_k &= \det\begin{pmatrix}\underline{y}_0 & \underline{y}_1-\underline{y}_0 & \underline{y}_2-\underline{y}_0 & \cdots & \underline{y}_n-\underline{y}_0\\ 1 & 0 & 0 & \cdots & 0\end{pmatrix}\\ &= (-1)^n\,\det\begin{pmatrix}\underline{y}_1-\underline{y}_0 & \underline{y}_2-\underline{y}_0 & \cdots & \underline{y}_n-\underline{y}_0\end{pmatrix}\\ &= (-1)^n\,\det Y,\end{aligned} \tag{3.5}$$

say, the first line being valid because the determinant of the matrix (1.4) remains the same if the first column is subtracted from the later ones, and the second line being elementary. We take this construction further by setting $\check{\underline{s}}_1=\underline{y}_1-\underline{y}_0$ and by forming the vectors

$$\check{\underline{s}}_i \;=\; (\underline{y}_i-\underline{y}_0)-\textstyle\sum_{j=1}^{i-1}\,\phi_{ij}\,(\underline{y}_j-\underline{y}_0), \qquad i=2,3,\ldots,n, \tag{3.6}$$

where the coefficients ϕ_{ij}, $1\leq j<i\leq n$, are given the values that minimize $\|\check{\underline{s}}_i\|$. Thus $\check{\underline{s}}_i$ is orthogonal to all vectors in the $(i-1)$-dimensional linear subspace of

$\mathcal{R}^n$ spanned by $\underline{y}_j-\underline{y}_0$, $j=1,2,\ldots,i-1$, which supplies $\underline{\check{s}}_i^T\underline{\check{s}}_j=0$, $1\le j<i\le n$. In other words, the $n\times n$ matrix $S=(\underline{\check{s}}_1\,\underline{\check{s}}_2\,\cdots\,\underline{\check{s}}_n)$ is derived by applying the Gram–Schmidt orthogonalization procedure to the columns of the matrix Y of equation (3.5). Because each column of S is the corresponding column of Y minus a linear combination of earlier columns, we have $\det S=\det Y$. Moreover, the mutual orthogonality of the columns of S implies that S^TS is the diagonal matrix with the diagonal elements $\|\underline{\check{s}}_i\|^2$, $i=1,2,\ldots,n$. Thus equation (3.5) gives the formula

$$|\det D_k| \;=\; |\det Y| \;=\; |\det S| \;=\; |\det(S^TS)|^{1/2} \;=\; \textstyle\prod_{i=1}^n \|\underline{\check{s}}_i\|. \tag{3.7}$$

Now the choice $\underline{\check{s}}_1=\underline{y}_1-\underline{y}_0$ with the coefficients that minimize $\|\underline{\check{s}}_i\|$ in expression (3.6) provide the bounds $\|\underline{\check{s}}_i\|\le|\underline{y}_i-\underline{y}_0\|$, $i=1,2,\ldots,n$. It follows from equation (3.7) that the assertion (3.3) is true.

Our justification of the condition (3.4) employs the vectors

$$\underline{\widehat{s}}_i \;=\; \underline{y}_i-\Big\{\underline{y}_0+\textstyle\sum_{j\in\{1,2,\ldots,n\}\backslash\{i\}}\psi_{ij}\,(\underline{y}_j-\underline{y}_0)\Big\}, \qquad i=1,2,\ldots,n, \tag{3.8}$$

the coefficients ψ_{ij} being given the values that minimize $\|\underline{\widehat{s}}_i\|$. The inequalities $\|\underline{\widehat{s}}_i\|\le\|\underline{\check{s}}_i\|$, $i=1,2,\ldots,n$, must hold, because all the freedom in the choice of $\underline{\check{s}}_i$ is available in the choice of $\underline{\widehat{s}}_i$. Moreover, the vector that occurs within the braces of expression (3.8) is a general point of the hyperplane that contains $\underline{y}_j$, $j\in\{0,1,\ldots,n\}\backslash\{i\}$, so $\|\underline{\widehat{s}}_i\|$ is the distance σ_i that has been defined already. These remarks imply $\sigma_i=\|\underline{\widehat{s}}_i\|\le\|\underline{\check{s}}_i\|$, $i=1,2,\ldots,n$. It follows from equation (3.7) that the assertion (3.4) is also true.

All the matrices D_k, $k=1,2,3,\ldots$, are required to be nonsingular, D_1 being given this property in the first paragraph of Section 2. Moreover, when the volume of the convex hull of the points $\underline{y}_i$, $i=0,1,\ldots,n$, is nonzero, the nonsingularity follows not only from the volume being $|\det D_k|/n!$ but also from inequality (3.4). Therefore the "alternative" steps preserve the nonsingularity of the interpolation matrix. The definition (1.4) shows that the elements of the new column of D_{k+1} are always the components of $\underline{x}_k+\underline{d}_k$ followed by a one, and we find in expression (2.8) that, after a "trust region" step, this new column is $D_k\underline{\theta}$, where $\underline{\theta}$ is the vector in $\mathcal{R}^{n+1}$ with the components θ_i, $i=0,1,\ldots,n$. Further, the replacement of the $(t+1)$-th column of D_k by $D_k\underline{\theta}$ is the same as replacing the whole matrix D_k by the product $D_k\Theta_t$, where Θ_t is the $(n+1)\times(n+1)$ identity matrix, except that its $(t+1)$-th column is $\underline{\theta}$. Thus we deduce from $\det(D_k\Theta_t)=\det D_k\det\Theta_k$ that D_{k+1} has the property

$$|\det D_{k+1}| \;=\; |\theta_t|\,|\det D_k|, \tag{3.9}$$

even in the case $F(\underline{x}_k+\underline{d}_k)<F(\underline{x}_k)$, because exchanging the first and $(t+1)$-th columns of D_{k+1} alters only the sign of $\det D_{k+1}$. Therefore, because condition (2.9) provides $|\theta_t|>0$, the "trust region" steps also preserve the nonsingularity of the interpolation matrix.

This paragraph is a diversion from our theory, in order to expose three strong advantages in practice of working with the $n \times n$ inverse matrix

$$Z \;=\; Y^{-1} \;=\; \left(\begin{array}{cccc} \underline{y}_1 - \underline{y}_0 & \underline{y}_2 - \underline{y}_0 & \cdots & \underline{y}_n - \underline{y}_0 \end{array} \right)^{-1}, \tag{3.10}$$

Y being nonsingular because of the nonsingularity of D_k in equation (3.5). We update Z when D_k is replaced by D_{k+1}, which can always be done in of magnitude n^2 computer operations. The first advantage occurs when $\underline{d}_k$ is an "alternative" step. Then we recall from the first complete paragraph after inequality (2.4) that $\underline{d}_k$ is a vector of length ρ that is orthogonal to the differences $\underline{y}_i - \underline{y}_0$, $i \in \{1, 2, \ldots, n\} \backslash \{t\}$. We see in equation (3.10) that this orthogonality is achieved by the t-th row of Z. Therefore the formula $\underline{d}_k = \pm \rho \, Z^T \underline{e}_t / \| Z^T \underline{e}_t \|$ provides the "alternative" step, where $\underline{e}_t$ is the t-th coordinate vector in $\mathcal{R}^n$. Secondly, all the distances σ_i, $i = 1, 2, \ldots, n$, are required when choosing the index t of an "alpha" step by applying the first part of expression (2.5). The distance σ_i from $\underline{y}_i$ to the hyperplane that contains the points $\underline{y}_\ell$, $\ell \in \{0, 1, \ldots, n\} \backslash \{i\}$ is $|\underline{v}_i^T (\underline{y}_i - \underline{y}_0)|$, where $\underline{v}_i$ is a vector of unit length that is orthogonal to the hyperplane, and where $\underline{y}_0$ can be replaced by any other point in the hyperplane. As before, the definition (3.10) shows that $\underline{v}_i^T$ is a multiple of the i-th row of Z. Therefore the required distances are given conveniently by the formula

$$\sigma_i \;=\; |\, \underline{e}_i^T Z (\underline{y}_i - \underline{y}_0) \,| \,/\, \| Z^T \underline{e}_i \| \;=\; 1 \,/\, \| Z^T \underline{e}_i \|, \qquad i = 1, 2, \ldots, n. \tag{3.11}$$

Thirdly, the parameters θ_i, $i = 1, 2, \ldots, n$, of expression (2.9) are easy to calculate when Z is available. Indeed, the equations (2.8) with $\underline{x}_k = \underline{y}_0$ show that $\underline{d}_k$ is the vector

$$\underline{d}_k \;=\; \textstyle\sum_{i=0}^n \theta_i \, (\underline{y}_i - \underline{y}_0) \;=\; \sum_{i=1}^n \theta_i \, (\underline{y}_i - \underline{y}_0), \tag{3.12}$$

where $\underline{d}_k$ is now a "trust region" step. It follows from equation (3.10) that the parameters θ_i, $i = 1, 2, \ldots, n$, are the components of the product $Z \underline{d}_k$.

Next we address the Lagrange functions Λ_j, $j = 0, 1, \ldots, n$, which are the linear polynomials from $\mathcal{R}^n$ to $\mathcal{R}$ that satisfy the conditions (3.1). For each j, we let $\Delta_j(\underline{x})$ be the $(n+1) \times (n+1)$ matrix that is formed by replacing $\underline{y}_j$ by $\underline{x}$ in the definition (1.4) of D_k, where $\underline{x}$ is a general point of $\mathcal{R}^n$. It follows that $\det \Delta_j(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, is a linear polynomial that satisfies $\Delta_j(\underline{y}_i) = 0$, $i \in \{0, 1, \ldots, n\} \backslash \{j\}$. Therefore the Lagrange functions are the ratios

$$\begin{aligned} \Lambda_j(\underline{x}) \;&=\; \det \Delta_j(\underline{x}) \,/\, \det \Delta_j(\underline{y}_j) \\ &=\; \det \Delta_j(\underline{x}) \,/\, \det D_k, \qquad \underline{x} \in \mathcal{R}^n, \qquad j = 0, 1, \ldots, n. \end{aligned} \tag{3.13}$$

A fundamental and well known property of these functions is that, if $\ell(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, is any constant or linear polynomial, then it can be expressed in the form

$$\ell(\underline{x}) \;=\; \textstyle\sum_{j=0}^n \ell(\underline{y}_j) \, \Lambda_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^n. \tag{3.14}$$

The assertion (3.2) is derived from equation (3.13) and from upper and lower bounds on $|\det \Delta_j(\underline{x})|$ and $|\det D_k|$, respectively. By regarding $\underline{x}$ as a new position of $\underline{y}_j$, inequality (3.3) gives the condition

$$|\det \Delta_j(\underline{x})| \;\leq\; \|\underline{x}-\underline{y}_0\| \prod_{i\in\{1,2,\ldots,n\}\setminus\{j\}} \|\underline{y}_i-\underline{y}_0\|, \qquad j=1,2,\ldots,n. \tag{3.15}$$

Therefore, if the assumptions (1.5) and (1.6) hold with $\rho_k=\rho$, equation (3.13) implies that the last n Lagrange functions have the property

$$|\Lambda_j(\underline{x})| \;\leq\; \|\underline{x}-\underline{y}_0\| \,(c_1\,\rho)^{n-1}/\,(c_2\,\rho^n), \qquad j=1,2,\ldots,n. \tag{3.16}$$

It follows that, if $\|\underline{x}-\underline{y}_0\|\leq\rho$, then $\sum_{j=1}^n |\Lambda_j(\underline{x})|$ is at most $n\,c_1^{n-1}/c_2$. Moreover, the choice $\ell(\underline{x})=1$, $\underline{x}\in\mathcal{R}^n$, in equation (3.14) provides $|\Lambda_0(\underline{x})|\leq 1+\sum_{j=1}^n |\Lambda_j(\underline{x})|$. Therefore the assertion (3.2) is true with $c_4=1+2\,n\,c_1^{n-1}/c_2$. We are now ready to establish the important inequality (1.7).

Lemma 1 If the interpolation points $\underline{y}_i$, $i=0,1,\ldots,n$, satisfy the conditions (1.5) and (1.6), with $\rho_k=\rho$ and $\underline{x}_k=\underline{y}_0$, where D_k is the matrix (1.4), and where c_1 and c_2 are positive constants, and if $\|\nabla^2 F\|$ and $\|\nabla^2 Q_k\|$ are uniformly bounded, where $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is a general twice differentiable function, and where $Q_k(\underline{x})$ $\underline{x}\in\mathcal{R}^n$, has the form (1.1), the vector $\underline{g}_k\in\mathcal{R}^n$ being defined by the equations (1.3) with $Q_k(\underline{y}_0)=F(\underline{y}_0)$, then the property (1.7) is achieved, where c_3 is another positive constant.

Proof The Lagrange functions Λ_j, $j=0,1,\ldots,n$, introduced in the first paragraph of this section, satisfy trivially the equation

$$\textstyle\sum_{j=0}^n \{F(\underline{y}_j)-Q_k(\underline{y}_j)\}\,\Lambda_j(\underline{x}) \;=\; 0, \qquad \underline{x}\in\mathcal{R}^n, \tag{3.17}$$

because of the interpolation conditions (1.3). We replace $F(\underline{y}_j)$ and $Q_k(\underline{y}_j)$ by their Taylor series expansions

$$\left.\begin{aligned} F(\underline{y}_j) &= F(\underline{y}_0)+(\underline{y}_j-\underline{y}_0)^T\underline{\nabla}F(\underline{y}_0)+\mathcal{O}(\|\underline{y}_j-\underline{y}_0\|^2) \\ Q_k(\underline{y}_j) &= Q_k(\underline{y}_0)+(\underline{y}_j-\underline{y}_0)^T\underline{\nabla}Q_k(\underline{y}_0)+\mathcal{O}(\|\underline{y}_j-\underline{y}_0\|^2) \end{aligned}\right\}, \tag{3.18}$$

where every $\mathcal{O}(\|\underline{y}_j-\underline{y}_0\|^2)$ term is a number of magnitude at most $\|\underline{y}_j-\underline{y}_0\|^2$, due to the uniformly bounded second derivatives. Inequality (1.5) allows these terms to be replaced by $\mathcal{O}(\rho^2)$. Thus, after cancelling the value $Q_k(\underline{y}_0)=F(\underline{y}_0)$, equations (3.17) and (3.18) provide the bound

$$\left|\textstyle\sum_{j=0}^n (\underline{y}_j-\underline{y}_0)^T\{\underline{\nabla}F(\underline{y}_0)-\underline{\nabla}Q_k(\underline{y}_0)\}\,\Lambda_j(\underline{x})\right| \;=\; \mathcal{O}(\rho^2)\,\textstyle\sum_{j=0}^n |\Lambda_j(\underline{x})|, \tag{3.19}$$

$\underline{x}\in\mathcal{R}^n$. Because the function $\ell(\underline{x})=(\underline{x}-\underline{y}_0)^T\{\underline{\nabla}F(\underline{y}_0)-\underline{\nabla}Q_k(\underline{y}_0)\}$, $\underline{x}\in\mathcal{R}^n$, is a linear polynomial, the identity (3.14) shows that the left hand side of expression (3.19) is just $|\ell(\underline{x})|$. We also recall that already we have deduced inequality (3.2) from the conditions (1.5) and (1.6). Therefore equation (3.19) gives the property

$$\left|(\underline{x}-\underline{y}_0)^T\{\underline{\nabla}F(\underline{y}_0)-\underline{\nabla}Q_k(\underline{y}_0)\}\right| \;=\; \mathcal{O}(\rho^2) \qquad \text{if} \qquad \|\underline{x}-\underline{y}_0\|\leq\rho. \tag{3.20}$$

Although we are regarding $\underline{x}-\underline{y}_0$ as a step in the space of the variables, there is no need to take this view. Instead, after dismissing the case $\underline{\nabla}F(\underline{y}_0)=\underline{\nabla}Q_k(\underline{y}_0)$ when inequality (1.7) is trivial, we let $\underline{x}-\underline{y}_0$ be the vector

$$\underline{x}-\underline{y}_0 \;=\; \{\underline{\nabla}F(\underline{y}_0)-\underline{\nabla}Q_k(\underline{y}_0)\}\,\rho\,/\,\|\underline{\nabla}F(\underline{y}_0)-\underline{\nabla}Q_k(\underline{y}_0)\| \qquad (3.21)$$

in equation (3.20). Thus we find that the bound (1.7) is true, which completes the proof. QED

4. Upper bounds on distances between interpolation points

We are going to show that the algorithms of Section 2 provide the property that the distances $\|\underline{y}_i-\underline{x}_k\|$, $i=1,2,\ldots,n$, are bounded above by a constant multiple of ρ on all iterations, which is needed in Lemma 1 above. The analysis includes the remark that, for every run of $3n+3$ consecutive iterations without a reduction in ρ, at least one of these iterations achieves a "successful" trust region step, where "successful" is defined immediately after expression (2.4). This remark is employed also in Sections 5 and 6, but no other details of the proof of Lemma 2 below are required later. Therefore we consider the frequency of "successful" trust region steps before presenting the formal statement of Lemma 2 and its justification. Thus, if the reader omits the proof of Lemma 2, which occupies most of this section, then the coherence of the paper is preserved.

Let p and q be any positive integers with $q\geq p+4$, such that no "successful" trust region steps and no reductions in ρ occur while the iteration number k satisfies $p\leq k\leq q$. The assertion in the previous paragraph, which we are going to prove next, is the bound $q\leq p+3n+1$. We recall from the penultimate paragraph of Section 2 that, as the "trust region" steps are assumed to be "unsuccessful", every three consecutive iterations with $p\leq k\leq q$ include at least one attempt at a "trust region" step and at least one attempt at a "beta" step. Therefore, if the iteration number k satisfies $p+3\leq k\leq q$, then the most recent previous attempt at a "trust region" step was "unsuccessful". It follows that, throughout these $q-p-2$ iterations, every attempt at a "beta" step is accepted, because otherwise the termination condition at the end of Section 2 would be achieved.

We complete the proof by considering the set $\mathcal{B}$, introduced in the paragraph that includes expression (2.6). The absence of "successful" trust region steps and reductions in ρ while the iteration number satisfies $p\leq k\leq q$ implies that $|\mathcal{B}|$ decreases monotonically during these iterations, where $|\mathcal{B}|$ is the number of elements of $\mathcal{B}$. Further, the value of $|\mathcal{B}|$ is at most $n-1$ at the beginning of the $(p+3)$-th iteration. Now $|\mathcal{B}|$ is reduced by one whenever a "beta" step is taken, so this happens at most $n-1$ times during the iterations with $p+3\leq k\leq q$. Also the number of steps $\underline{d}_k$ that are not "beta" steps during these iterations is at most $2n$, because we have noted already that every three consecutive iterations include at least one "beta" step. Thus $q-p-2$, which is the total number of iterations with $p+3\leq k\leq q$, is at most $3n-1$. This conclusion is exactly the required bound $q\leq p+3n+1$.

Having proved that every run of $3n+3$ consecutive iterations without a reduction in ρ includes at least one "successful" trust region step, we now present the rest of the argument that establishes condition (1.5).

Lemma 2 The algorithms of Section 2 provide the property that, on every iteration, the interpolation points satisfy the conditions

$$\|\underline{y}_i - \underline{x}_k\| \;=\; \|\underline{y}_i - \underline{y}_0\| \;\leq\; c_1\,\rho, \qquad i=1,2,\ldots,n, \tag{4.1}$$

where c_1 is a positive constant and ρ is the trust region radius.

Proof Let $\underline{y}_i$ and $\underline{y}_i^+$, $i=0,1,\ldots,n$, be the interpolation points at the beginning and end of the k-th iteration, respectively. We recall from the penultimate paragraph of Section 1 that only $\underline{y}_t$ and $\underline{y}_0$ may be changed during the iteration, $\underline{y}_t$ being replaced by $\underline{x}_k+\underline{d}_k$, and then $\underline{y}_0$ being switched with the new $\underline{y}_t$ if and only if the reduction $F(\underline{x}_k+\underline{d}_k)<F(\underline{x}_k)$ is achieved. It follows from $\underline{y}_0=\underline{x}_k$ and $\|\underline{d}_k\|\leq\rho$ that this construction gives the bounds

$$\|\underline{y}_t^+ - \underline{y}_0^+\| \;\leq\; \rho \quad\text{and}\quad \|\underline{y}_i^+ - \underline{y}_0^+\| \;\leq\; \|\underline{y}_i - \underline{y}_0\| + \rho, \quad i=1,2,\ldots,n. \tag{4.2}$$

We combine them with the operations on the set $\mathcal{B}$ mentioned above. Let i be any integer from $[1,n]$ that is not an element of $\mathcal{B}$ at the beginning of the k-th iteration. Then $\underline{y}_i$ was the new interpolation point $\underline{x}_\ell+\underline{d}_\ell$ of the ℓ-th iteration, for some integer $\ell<k$ that is greater than the number of the most recent iteration that picked $\mathcal{B}=\{1,2,\ldots,n\}$. This choice of $\mathcal{B}$ occurs for each new value of ρ and immediately after every "successful" trust region step. Therefore the analysis in the second and third paragraphs of this section supplies $\ell\geq k-3n-2$. We let ℓ be as large as possible, in order that $\underline{y}_i^+=\underline{y}_i$ holds on all iterations with numbers $\ell+1,\ell+2,\ldots,k-1$. Further, the value of $\|\underline{y}_i-\underline{y}_0\|$ is no greater than ρ at the beginning of the $(\ell+1)$-th iteration, and each iteration with a number in the interval $[\ell+1,k-1]$ increases $\|\underline{y}_i-\underline{y}_0\|$ by at most ρ, due to the bounds (4.2). Thus, on the k-th iteration for general k, we deduce the property

$$\|\underline{y}_i - \underline{y}_0\| \;\leq\; (k-\ell)\,\rho \;\leq\; (3n+2)\,\rho, \qquad i\in\{1,2,\ldots,n\}\backslash\mathcal{B}. \tag{4.3}$$

At the beginning of the first iteration, we may regard ρ and $\|\underline{y}_i-\underline{y}_0\|$, $i=1,2,\ldots,n$, as constants. Therefore condition (4.1) is achieved initially by making c_1 sufficiently large. Moreover, we recall from the last paragraph of Section 2 that the iterations with the current value of ρ are complete only if $\|\underline{y}_i-\underline{y}_0\|\leq\beta\rho$, $i\in\mathcal{B}$, holds, and then inequality (4.3) gives $\|\underline{y}_i-\underline{y}_0\|\leq\max[3n+2,\beta]\,\rho$, $i=1,2,\ldots,n$. We also recall from the second paragraph of Section 2 that every reduction in ρ is by a factor of 10. It follows that the conditions

$$\|\underline{y}_i - \underline{y}_0\| \;\leq\; 10\,\max[3n+2,\beta]\,\rho, \qquad i=1,2,\ldots,n, \tag{4.4}$$

are satisfied immediately after each change to ρ, so we require the value of c_1 to be at least $10\max[3n+2,\beta]$. It remains to prove that the lemma is true during any sequence of iterations that does not alter ρ.

We define Γ_k to be the sum $\sum_{i=1}^n \|\underline{y}_i - \underline{y}_0\|$ on the k-th iteration for every k. We are going to prove that Γ_k is bounded above by a constant multiple of ρ. Expression (4.2) shows that $\|\underline{y}_t^+ - \underline{y}_0^+\|$ is substantially less than $\|\underline{y}_t - \underline{y}_0\|$ if $\|\underline{y}_t - \underline{y}_0\|$ is sufficiently large, which is usual when a "beta" step is taken. Thus a "beta" step can cause the new sum $\Gamma_{k+1} = \sum_{i=1}^n \|\underline{y}_i^+ - \underline{y}_0^+\|$ to be less than Γ_k by a large multiple of ρ. Such reductions can compensate for any increases in the sum on the other iterations, these increases being bounded by the condition

$$\Gamma_{j+1} \;\leq\; \Gamma_j + n\rho, \tag{4.5}$$

which is a direct consequence of the inequalities (4.2), where j is the number of any iteration that is given the current value of ρ.

The purpose of the parameter τ_β is to provide enough "beta" steps for our proof of convergence. We recall from the sentence after inequality (2.6) that at most τ_β "trust region" steps are taken without an attempt at a "beta" step. Moreover, we recall from the penultimate paragraph of Section 2 that, if an attempt at a "trust region" step is "unsuccessful", then a "beta" step is tried before the next attempt at a "trust region" step. Therefore, between any two consecutive attempts at "beta" steps, there are at most τ_β attempts at "trust region" steps; also some "alpha" steps may be tried, the number of trials being at most $\tau_\beta + 1$, because "alpha" step attempts are separated by "trust region" step attempts. It follows that the number of consecutive iterations without trying a "beta" step is bounded above by $2\tau_\beta + 1$.

Let k be the number of an iteration that attempts a "beta" step, and assume for the moment that Γ_k has the lower bound

$$\Gamma_k \;>\; n \,\max\left[\, 3n+2,\, \beta,\, (2\tau_\beta+2)\, n \,\right] \rho \;=\; c_5\, \rho, \tag{4.6}$$

say. The definition $\Gamma_k = \sum_{i=1}^n \|\underline{y}_i - \underline{y}_0\|$ with the assumption (4.6) imply $\|\underline{y}_t - \underline{y}_0\| > (3n+2)\rho$ and $\|\underline{y}_t - \underline{y}_0\| > \beta\rho$, where $\|\underline{y}_t - \underline{y}_0\|$ is the greatest of the distances $\|\underline{y}_i - \underline{y}_0\|$, $i = 1, 2, \ldots, n$. It follows from condition (4.3) that t is an element of $\mathcal{B}$, so $\|\underline{y}_t - \underline{y}_0\|$ is also the greatest of the distances $\|\underline{y}_i - \underline{y}_0\|$, $i \in \mathcal{B}$. Therefore both parts of expression (2.6) are satisfied, which makes the attempt at a "beta" step successful on the k-th iteration.

This "beta" step provides the property

$$\begin{aligned} \Gamma_{k+1} \;&\leq\; \Gamma_k + n\rho - \|\underline{y}_t - \underline{y}_0\| \\ &\leq\; \Gamma_k + n\rho - n^{-1}\,\Gamma_k \;<\; \Gamma_k - (2\tau_\beta+1)\, n\rho, \end{aligned} \tag{4.7}$$

the first line being due to the bounds (4.2), and the second line being due to the choice of t and to the assumption (4.6). Let ℓ be the number of the next iteration that includes an attempt at a "beta" step. We found that ℓ is at most $k + 2\tau_\beta + 2$ in the paragraph before last. Thus inequalities (4.7) and (4.5) give the conditions

$$\Gamma_j \;<\; \Gamma_k, \qquad j = k+1, k+2, \ldots, \ell. \tag{4.8}$$

On the other hand, if k and ℓ are the numbers of iterations that make consecutive attempts at "beta" steps, and if the bound (4.6) fails, then inequality (4.5) with $\ell \leq k+2\tau_\beta+2$ supply the condition

$$\Gamma_j \;\leq\; \{c_5 + (2\tau_\beta+2)\, n\}\, \rho, \qquad j=k+1, k+2, \ldots, \ell. \tag{4.9}$$

We recall from the second paragraph of Section 2 that κ is the number of the first iteration that employs the current value of ρ, and the remarks in the first two paragraphs of this proof provide $\Gamma_\kappa \leq c_6\rho$, where c_6 is another positive constant. Hence, corresponding to the conditions (4.9), the inequalities

$$\Gamma_j \;\leq\; \{c_6 + (2\tau_\beta+2)\, n\}\, \rho, \qquad j=\kappa, \kappa+1, \ldots, \widehat{\kappa}, \tag{4.10}$$

hold, where $\widehat{\kappa}$ is the number of the first iteration that attempts a "beta" step with the current ρ.

Let j be the number of any iteration that is given the current value of ρ. Because we are seeking an upper bound on Γ_j that is valid for every j, we may assume without loss of generality that Γ_j is the largest of the numbers Γ_i, $i = \kappa, \kappa+1, \ldots, j$. We employ expression (4.10) in the case $j \leq \widehat{\kappa}$. Otherwise, we let k and ℓ be the numbers of the iterations that make consecutive attempts at "beta" steps with $\widehat{\kappa} \leq k < j \leq \ell$. The value of ρ given to the ℓ-th iteration is always the current one, because it is stated at the end of Section 2 that a "beta" step is attempted on the last iteration with the current ρ. Our assumption includes $\Gamma_k \leq \Gamma_j$, which rules out the possibility (4.8). It follows that inequality (4.6) fails, the alternative being that condition (4.9) is satisfied. Therefore, for every j under consideration, the property

$$\textstyle\sum_{i=1}^n \|\underline{y}_i - \underline{y}_0\| \;=\; \Gamma_j \;\leq\; \{\max[c_5, c_6] + (2\tau_\beta+2)\, n\}\, \rho \tag{4.11}$$

is achieved, which establishes that all the distances $\|\underline{y}_i - \underline{y}_0\|$, $i = 1, 2, \ldots, n$, are bounded above by a constant multiple of ρ. The proof is complete. QED

5. Lower bounds on determinants of interpolation matrices

Inequality (1.6) is justified in this section, using the bound (3.4) and Lemma 2. Again the reader may skip the proof without losing the coherence of the paper.

Lemma 3 The algorithms of Section 2 provide the property that, on every iteration, the determinant of the interpolation matrix (1.4) satisfies the condition

$$|\det D_k| \;\geq\; c_2\, \rho^n, \tag{5.1}$$

where c_2 is a positive constant and ρ is the trust region radius.

Proof The paragraph that includes expression (2.5) states that at most τ_α "trust region" steps may be taken for the current ρ without an attempt at an "alpha"

step. Moreover, we found at the beginning of Section 4 that every run of $3n+3$ consecutive iterations with the current ρ includes at least one "successful" trust region step. Therefore we may let $\widehat{\tau}_\alpha$ be a constant integer such that every run of $\widehat{\tau}_\alpha$ consecutive iterations with the current ρ includes at least one attempt at an "alpha" step.

We require positive lower bounds on the ratio $|\det D_{k+1}|/|\det D_k|$ in the three cases when the k-th iteration takes an "alpha" or a "beta" or a "trust region" step. The identity (3.9) is satisfied for a "trust region" step, θ_t being given by expressions (2.8) and (2.9), and we recall that the equations (2.8) with $\underline{y}_0=\underline{x}_k$ supply the formula (3.12). Therefore the choice (2.9) with Lemma 2 yield the property

$$\|\underline{d}_k\| \;=\; \|\,\textstyle\sum_{i=1}^n\theta_i\,(\underline{y}_i-\underline{y}_0)\,\| \;\le\; |\theta_t|\,\textstyle\sum_{i=1}^n\|\underline{y}_i-\underline{y}_0\| \;\le\; n\,c_1\,|\theta_t|\,\rho. \tag{5.2}$$

Thus the bound

$$|\det D_{k+1}| \;=\; |\theta_t|\,|\det D_k| \;\ge\; |\det D_k|\,\|\underline{d}_k\|/(nc_1\rho) \;\ge\; |\det D_k|/(2nc_1) \tag{5.3}$$

is achieved in the "trust region" case, the last part being due to the second of the necessary conditions (2.7) for a "trust region" step.

We treat the "alpha" and "beta" cases by employing the remark, given after the definition (1.4), that $|\det D_k|/n!$ is the volume of the convex hull of the points $\underline{y}_i\in\mathcal{R}^n$, $i=0,1,\ldots,n$. Indeed, because the k-th iteration replaces $\underline{y}_t$ by $\underline{x}_k+\underline{d}_k=\underline{y}_0+\underline{d}_k$, the ratio $|\det D_{k+1}|/|\det D_k|$ is just σ_t^+/σ_t, where σ_t and σ_t^+ are the distances from $\underline{y}_t$ and $\underline{y}_0+\underline{d}_k$, respectively, to the hyperplane that contains the points $\underline{y}_j$, $j\in\{0,1,\ldots,n\}\backslash\{t\}$. In both cases the step $\underline{d}_k$ is the "alternative" one, defined in the paragraph after expression (2.4), and having the property $\sigma_t^+=\|\underline{d}_k\|=\rho$. Moreover, in the "alpha" case σ_t is the least of the positive numbers σ_i, $i=1,2,\ldots,n$, that satisfy inequality (3.4), which implies $\sigma_t\le|\det D_k|^{1/n}$, and it is sufficient in the "beta" case that Lemma 2 provides $\sigma_t\le\|\underline{y}_t-\underline{y}_0\|\le c_1\rho$. Thus we deduce the bound

$$|\det D_{k+1}| \;=\; (\sigma_t^+/\sigma_t)\,|\det D_k| \;\ge\; \rho\,|\det D_k|/|\det D_k|^{1/n} \tag{5.4}$$

or

$$|\det D_{k+1}| \;=\; (\sigma_t^+/\sigma_t)\,|\det D_k| \;\ge\; |\det D_k|/c_1, \tag{5.5}$$

when the k-th iteration takes an "alpha" step or a "beta" step, respectively.

Let k be the number of an iteration that attempts an "alpha" step, and let ℓ be the number of the next iteration that also attempts an "alpha" step. We consider the value of $|\det D_{j+1}|$ when the iteration number j is in the interval $[k,\ell-1]$. The value of ρ is constant during these iterations, because the calculations with each new value of ρ begin by attempting an "alpha" step, as stated in the penultimate paragraph of Section 2. Furthermore, every iteration number j is in a unique interval $[k,\ell-1]$, where k and ℓ are numbers of iterations that make consecutive

attempts at "alpha" steps. The definition of $\widehat{\tau}_\alpha$ in the first paragraph of this proof provides $\ell \le k+\widehat{\tau}_\alpha$.

Because either a "trust region" or a "beta" step is taken on the iterations under consideration, expressions (5.3) and (5.5) give the bounds

$$|\det D_{j+1}| \ \ge \ |\det D_j| \,/\, (2nc_1), \qquad j=k+1, k+2, \ldots, \ell-1. \tag{5.6}$$

We are also going to establish that the inequalities

$$|\det D_{j+1}| \ > \ |\det D_k|, \qquad j=k, k+1, \ldots, \ell-1, \tag{5.7}$$

are achieved if $|\det D_k|$ is sufficiently small. Therefore for the moment we make the assumption

$$|\det D_k| \ < \ \{\min[\alpha, (2nc_1)^{-\widehat{\tau}_\alpha}]\}^n \rho^n \ = \ c_7 \rho^n, \tag{5.8}$$

say. By combining the α term of this assumption with the bound (3.4), we find that $\prod_{i=1}^n \sigma_i \le |\det D_k| < (\alpha\rho)^n$ holds on the k-th iteration. Hence, as stated in the paragraph that includes expression (2.5), this iteration actually takes an "alpha" step. Thus, because of inequality (5.4), condition (5.8) provides the property

$$|\det D_{k+1}| \ \ge \ \rho\,|\det D_k| \,/\, |\det D_k|^{1/n} \ > \ (2nc_1)^{\widehat{\tau}_\alpha}\,|\det D_k|. \tag{5.9}$$

It follows from expression (5.6) with $\ell \le k+\widehat{\tau}_\alpha$ that the bounds (5.7) are satisfied as required.

Whenever the k-th iteration takes an "alpha" step, the volume of the convex hull of the interpolation points is increased, which is the condition $|\det D_{k+1}| > |\det D_k|$, the alternative being that a "trust region" or "beta" step is taken, giving the weaker bound (5.6) with $j=k$. Therefore, if k and ℓ are indices of iterations that make consecutive attempts at "alpha" steps as before, but if assumption (5.8) fails, then the conditions (5.6) with $\ell \le k+\widehat{\tau}_\alpha$ supply the inequalities

$$|\det D_{j+1}| \ \ge \ |\det D_k| \,/\, (2nc_1)^{\widehat{\tau}_\alpha} \ \ge \ c_7 \rho^n / \,(2nc_1)^{\widehat{\tau}_\alpha}, \qquad k \le j < \ell. \tag{5.10}$$

Expressions (5.7) or (5.10) are valid when the assumption (5.8) holds or fails, respectively. Together they yield the bound

$$|\det D_{j+1}| \ \ge \ \min[\,|\det D_k|, \, c_7 \rho^n / \,(2n\,c_1)^{\widehat{\tau}_\alpha}], \qquad k \le j < \ell, \tag{5.11}$$

for every value of $|\det D_k|$.

The property (5.11) is the inequality

$$|\det D_{j+1}| \ \ge \ \min[\,|\det D_\kappa|, \, c_7 \rho_\kappa^n / \,(2n\,c_1)^{\widehat{\tau}_\alpha}] \tag{5.12}$$

in the case $k=\kappa$, the notation ρ_κ being used for ρ because it is useful later, where κ is still the number of the first iteration that employs the current value of ρ. The following argument shows that the bound (5.12) remains true when

the j-th iteration employs $\rho=\rho_\kappa$, but k is greater than κ in expression (5.11). If this assertion is false, we let j be the least integer such that failure occurs, which implies $|\det D_{j+1}|<|\det D_{i+1}|$, $i=\kappa,\kappa+1,\ldots,j-1$, because the right hand side of expression (5.12) is independent of j. Thus $|\det D_{j+1}|$ is less than $|\det D_k|$ in condition (5.11), so this condition reduces to $|\det D_{j+1}|\geq c_7\rho_\kappa^n/(2nc_1)^{\widehat{\tau}_\alpha}$, giving the contradiction that inequality (5.12) is satisfied as required.

We complete the proof by letting c_2 be the constant

$$c_2 \;=\; \min\left[\,|\det D_1|\,/\,\rho_1^n,\; c_7\,/\,(2nc_1)^{\widehat{\tau}_\alpha}\,\right], \tag{5.13}$$

and by showing that the bound (5.1) is achieved for every iteration number k. The choice (5.13) implies $c_2\leq|\det D_1|\,/\rho_1^n$, which covers the case $k=1$. Furthermore, if j is the number of any iteration that picks the new point $\underline{x}_j+\underline{d}_j$ using the initial trust region radius $\rho=\rho_1$, then inequality (5.12) is satisfied with $\kappa=1$, which gives the alleged property

$$|\det D_{j+1}| \;\geq\; \min\left[\,|\det D_1|,\; c_7\,\rho_1^n/\,(2n\,c_1)^{\widehat{\tau}_\alpha}\,\right] \;=\; c_2\,\rho^n. \tag{5.14}$$

Therefore it is sufficient to establish that, if c_2 is the constant (5.13), and if the condition (5.1) holds at the beginning of every iteration that is given the current value of ρ, then this property is also achieved for the next value of ρ. The proof is completed by induction.

We continue to let κ be the number of an iteration that reduces the trust region radius, the values of ρ at the end and start of this iteration being ρ_κ and $10\rho_\kappa$, respectively. In accordance with the remarks at the end of the previous paragraph, we assume that condition (5.1) holds on every iteration that is given the trust region radius $10\rho_\kappa$, which supplies $|\det D_\kappa|\geq c_2(10\rho_\kappa)^n>c_2\rho_\kappa^n$. By substituting this bound into inequality (5.12) we find that the new matrices D_{j+1}, generated by the iterations that employ $\rho=\rho_\kappa$, have the property

$$|\det D_{j+1}| \;\geq\; \min\left[\,c_2\,\rho_\kappa^n,\; c_7\,\rho_\kappa^n\,/\,(2nc_1)^{\widehat{\tau}_\alpha}\,\right] \;=\; c_2\,\rho_\kappa^n, \tag{5.15}$$

the last assertion being valid because the number (5.13) satisfies $c_2\leq c_7/(2nc_1)^{\widehat{\tau}_\alpha}$. Therefore the lemma is true. QED

6. Weak convergence

The bounded second derivatives of the objective function F and the quadratic models Q_k are important to our analysis of convergence. We let the constants Φ and Ω be upper bounds on $\|\nabla^2F(\underline{x})\|$, $\underline{x}\in\mathcal{R}^n$, and $\|G_k\|=\|\nabla^2Q_k\|$, $k=1,2,3,\ldots$, respectively. The range $\underline{x}\in\mathcal{R}^n$ can be replaced by the union of the trust regions $\{\underline{x}:\|\underline{x}-\underline{x}_k\|\leq\rho_k\}$, $k=1,2,3,\ldots$. Another important assumption is that F is bounded below.

Weak convergence is proved in this section, which means that the algorithms of Section 2 have the property that the gradient norms $\|\underline{\nabla}F(\underline{x}_k)\|$, $k=1,2,3,\ldots$,

are not bounded away from zero. There are two separate parts of the analysis, namely showing the termination of the sequence of iterations with each value of ρ, and establishing the inequality $\|\underline{\nabla} F(\underline{x}_k)\| \le c_8 \rho$ on at least one iteration with the current value of ρ, where c_8 is another positive constant. The first part gives the limit $\rho_k \to 0$ as $k \to \infty$, all changes to ρ being reductions by a factor of 10, and then the bounds $\|\underline{\nabla} F(\underline{x}_k)\| \le c_8 \rho_k$, for suitable values of k, establish that weak convergence is achieved.

In order to retain the structure of the paper, where proofs of lemmas can be omitted without loss of coherence, there is only one lemma in this section, given at the end, which shows termination of the iterations with the current value of ρ. We address first the bound in the previous paragraph that employs the constant c_8. We pick the value

$$c_8 \;=\; (\tfrac{19}{9}\, c_3 + \tfrac{5}{4}\, \Omega + \tfrac{5}{9}\, \Phi)\; \max\,[\gamma, 1], \tag{6.1}$$

the constants c_3 and γ being taken from expressions (1.7) and (2.7), respectively. It is proved in the next three paragraphs that this choice supplies the following property. If the bound

$$\|\underline{\nabla} F(\underline{x}_k)\| \;>\; c_8\, \rho \tag{6.2}$$

holds, and if the k-th iteration tries to take a "trust region" step, then all the conditions (2.7) and (2.4) are achieved, which makes the attempt "successful".

We assume inequality (6.2) and that $\underline{d}_k$ is a "trust region" step, so it has to satisfy the conditions (2.2). Expression (1.1) with the definition of the "Cauchy" step $\widehat{\underline{d}}_k$ imply the bound

$$\begin{aligned} Q_k(\underline{x}_k + \widehat{\underline{d}}_k) \;&\le\; Q_k\Big(\underline{x}_k - \rho\, \underline{g}_k \,/\, \|\underline{g}_k\|\Big) \;\le\; Q_k(\underline{x}_k) - \rho\, \|\underline{g}_k\| + \tfrac{1}{2}\, \Omega\, \rho^2 \\ &\le\; Q_k(\underline{x}_k) - \tfrac{1}{2}\, \rho\, \Big(\|\underline{g}_k\| + \|\underline{\nabla} F(\underline{x}_k)\|\Big) + \tfrac{1}{2}\, (c_3 + \Omega)\, \rho^2, \end{aligned} \tag{6.3}$$

the last line being due to Lemma 1. It follows from the relations (2.2), (6.2) and (6.1) that $\underline{d}_k$ has the property

$$\begin{aligned} Q_k(\underline{x}_k + \underline{d}_k) \;\le\; Q_k(\underline{x}_k + \widehat{\underline{d}}_k) \;&\le\; Q_k(\underline{x}_k) - \tfrac{1}{2}\, \rho\, \|\underline{g}_k\| + \tfrac{1}{2}\, \rho^2\, (-c_8 + c_3 + \Omega) \\ &\le\; Q_k(\underline{x}_k) - \tfrac{1}{2}\, \rho\, \|\underline{g}_k\| - \tfrac{1}{8}\, \Omega\, \rho^2. \end{aligned} \tag{6.4}$$

On the other hand, if $\underline{d}$ is any vector with $\|\underline{d}\| < \frac{1}{2}\rho$, then expression (1.1) shows that $Q_k(\underline{x}_k + \underline{d})$ is strictly greater than the right hand side of inequality (6.4), which excludes $\underline{d} = \underline{d}_k$. Therefore the bound $\|\underline{d}_k\| \ge \frac{1}{2}\rho$ is achieved, which is the second of the conditions (2.7) that are necessary for the "success" of an attempt at a "trust region" step.

Next we require a bound on $|Q_j(\underline{x}_j + \underline{d}_j) - F(\underline{x}_j + \underline{d}_j)|$, where j is the number of any iteration that employs the current value of ρ. The constants Ω and Φ in the first paragraph of this section, with $\|\underline{d}_j\| \le \rho$, provide the properties

$$\left.\begin{aligned} |\, Q_j(\underline{x}_j + \underline{d}_j) - Q_j(\underline{x}_j) - \underline{d}_j^T \underline{\nabla} Q_j(\underline{x}_j)\,| \;&\le\; \tfrac{1}{2}\, \Omega\, \rho^2 \\ |\, F(\underline{x}_j + \underline{d}_j) - F(\underline{x}_j) - \underline{d}_j^T \underline{\nabla} F(\underline{x}_j)\,| \;&\le\; \tfrac{1}{2}\, \Phi\, \rho^2 \end{aligned}\right\}. \tag{6.5}$$

Thus the identity $Q_j(\underline{x}_j)=F(\underline{x}_j)$ and Lemma 1 supply the condition

$$|\,Q_j(\underline{x}_j+\underline{d}_j)-F(\underline{x}_j+\underline{d}_j)\,|\;\leq\;(c_3+\tfrac{1}{2}\,\Omega+\tfrac{1}{2}\,\Phi)\,\rho^2, \tag{6.6}$$

which is used in two ways. Firstly, because of the definition (2.3), the right hand side $\gamma\eta_k$ of the first part of expression (2.7) is no greater than $(c_3+\frac{1}{2}\Omega+\frac{1}{2}\Phi)\gamma\rho^2$, while the left hand side is bounded below by the inequality

$$\begin{aligned}Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)\;&\geq\;Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\widehat{\underline{d}}_k)\;\geq\;\rho\,\|\underline{g}_k\|-\tfrac{1}{2}\,\Omega\,\rho^2\\&\geq\;\rho\,\|\underline{\nabla}F(\underline{x}_k)\|-(c_3+\tfrac{1}{2}\,\Omega)\,\rho^2\;\geq\;(c_8-c_3-\tfrac{1}{2}\,\Omega)\,\rho^2,\end{aligned} \tag{6.7}$$

which is deduced from the conditions (2.2) on the "trust region" step $\underline{d}_k$, from the first line of expression (6.3), from Lemma 1, and from the assumption (6.2). It follows that the first of the conditions (2.7) for a "successful" trust region step is achieved as required if we pick a value of c_8 that makes $(c_8-c_3-\frac{1}{2}\Omega)$ greater than $(c_3+\frac{1}{2}\Omega+\frac{1}{2}\Phi)\gamma$. We see that the choice (6.1) is adequate.

Therefore, when $\|\underline{\nabla}F(\underline{x}_k)\|>c_8\rho$ holds and when the k-th iteration generates a "trust region" step $\underline{d}_k$, then, as explained in the paragraph that includes expression (2.7), the new function value $F(\underline{x}_k+\underline{d}_k)$ is calculated. It remains to show in this case that the condition (2.4) for the "success" of the step is satisfied too. Because $F(\underline{x}_k)$ and $Q_k(\underline{x}_k)$ are the same, this condition is equivalent to the requirement

$$\begin{aligned}0.9\,\{Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)\}\;&\geq\;Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)-\{F(\underline{x}_k)-F(\underline{x}_k+\underline{d}_k)\}\\&=\;F(\underline{x}_k+\underline{d}_k)-Q_k(\underline{x}_k+\underline{d}_k).\end{aligned} \tag{6.8}$$

The properties (6.7) and (6.6) show that the left and right hand sides of expression (6.8) are bounded below and above by $0.9(c_8-c_3-\frac{1}{2}\Omega)\rho^2$ and by $(c_3+\frac{1}{2}\Omega+\frac{1}{2}\Phi)\rho^2$, respectively. Therefore, in order to satisfy inequality (2.4), our final condition on c_8 is the constraint

$$0.9\,(c_8-c_3-\tfrac{1}{2}\,\Omega)\;\geq\;c_3+\tfrac{1}{2}\,\Omega+\tfrac{1}{2}\,\Phi\quad\Longleftrightarrow\quad c_8\;\geq\;\tfrac{19}{9}\,c_3+\tfrac{19}{18}\,\Omega+\tfrac{5}{9}\,\Phi, \tag{6.9}$$

which is also achieved by the choice (6.1). It follows from the conclusions of the last three paragraphs that, if the k-th iteration tries to take a "trust region" step, and if the attempt is not "successful", then $\|\underline{\nabla}F(\underline{x}_k)\|$ is at most $c_8\rho$.

We recall from the last paragraph of Section 2 that the iterations with the current value of ρ are terminated only after an unsuccessful attempt at a "trust region" step. Then the analysis above provides the bound $\|\underline{\nabla}F(\underline{x}_k)\|\leq c_8\rho$ on at least one iteration with the current ρ. Thus weak convergence is achieved, as mentioned already, provided ρ tends to zero. In other words, weak convergence occurs if the sequence of iterations with any constant value of ρ is finite, which is proved below in Lemma 4, after a remark on the "exact" trust region steps of the algorithms.

It is stated in Section 2 that, if the number (2.3) is zero for sufficiently large k, then the "trust region" step $\underline{d}_k$ is required to minimize $Q_k(\underline{x}_k+\underline{d}_k)$ subject

to $\|\underline{d}_k\| \leq \rho$, instead of being any vector that satisfies the conditions (2.2). The reason for the stronger conditions on $\underline{d}_k$ when η_k is zero is shown by the following example. Let F be the quadratic function

$$F(\underline{x}) \;=\; \xi^2 + 3\,\eta^2 + 0\,\zeta^2, \qquad \underline{x} \in \mathcal{R}^3, \tag{6.10}$$

where ξ, η and ζ are the components of $\underline{x} \in \mathcal{R}^3$, the factor 0 making F independent of ζ, let the initial trust region radius be $\rho = 2$, let the initial quadratic model be $Q_1 \equiv F$, and let $\underline{x}_1$ be so close to the origin that every "Cauchy" step $\widehat{\underline{d}}_k$ satisfies $\|\widehat{\underline{d}}_k\| \leq 1$. Because $Q_k \equiv F$ allows $Q_{k+1} \equiv Q_k$ due to $Q_k(\underline{x}_k + \underline{d}_k) = F(\underline{x}_k + \underline{d}_k)$, we find by induction that $Q_k \equiv F$ can hold for every $k \geq 1$. Thus all the numbers (2.3) are zero. The purpose of the example is to show that the number of iterations with the initial ρ may be infinite if "trust region" steps do not have to be "exact". Indeed, we let every "trust region" step be the sum $\underline{d}_k = \widehat{\underline{d}}_k + \underline{e}_3$, where $\widehat{\underline{d}}_k$ is still the "Cauchy" step, and where $\underline{e}_3$ is the third coordinate vector in $\mathcal{R}^3$. This choice has the properties $Q_k(\underline{x}_k + \underline{d}_k) = Q_k(\underline{x}_k + \widehat{\underline{d}}_k)$, $\|\underline{d}_k\| \leq \rho$ and $\|\underline{d}_k\| > \frac{1}{2}\rho$, so all the inequalities (2.2) and (2.7) are achieved, the value of η_k being zero. Therefore the new function value $F(\underline{x}_k + \underline{d}_k)$ is calculated. The identity $Q_k \equiv F$ implies that $F(\underline{x}_k) - F(\underline{x}_k + \underline{d}_k)$ is the positive number $Q_k(\underline{x}_k) - Q_k(\underline{x}_k + \underline{d}_k)$. It follows from condition (2.4) that every attempt at a "trust region" step is "successful". Thus the conditions for terminating the sequence of iterations, given in the last paragraph of Section 2, are never achieved. The particular choice (6.10) of the objective function has the property that, if $\underline{x}_1$ has the components $\xi_1 = 1$, $\eta_1 = 1/3$ and $\zeta_1 = 0$, and if the "alpha" and "beta" steps make no changes to the centre of the trust region,then, for every positive integer j, the point $\underline{x}_{k+1} = \underline{x}_k + \underline{d}_k$ of the j-th "trust region" step has the components $\xi_{k+1} = 2^{-j}$, $\eta_{k+1} = (-2)^{-j}/3$ and $\zeta_{k+1} = j$.

Lemma 4 Let any algorithm from Section 2 be applied to a function F that, as usual, is bounded below and has bounded second derivatives. The number of iterations with each value of ρ is finite.

Proof The difference $F(\underline{x}_k) - F(\underline{x}_{k+1})$ tends to zero as $k \to \infty$, because the sequence $F(\underline{x}_k)$, $k = 1, 2, 3, \ldots$, is monotonically decreasing and bounded below, but we are going to find in several situations that every "successful" trust region step $\underline{d}_k$ with the current value of ρ has the property $F(\underline{x}_k) - F(\underline{x}_k + \underline{d}_k) \geq \widehat{c}$, where $\widehat{c}$ is a positive number that depends on ρ but not on k. Then it follows from equation (1.2) that the number of "successful" trust region steps with the current ρ is finite. Moreover, we recall from the beginning of Section 4 that at least one "successful" trust region step occurs in every run of $3n+3$ consecutive iterations with fixed ρ. In those situations, therefore, the sequence of iterations with the current ρ does terminate as required.

One of the situations occurs when the number (2.3) becomes positive during the iterations with the current ρ. We let $\eta_{\check{\kappa}}$ be the first positive value of η_k, which is a lower bound on later values of η_k, and we assume that the k-th iteration takes a "successful" trust region step, where $k \geq \check{\kappa}$. Because the conditions (2.4) and

(2.7) are necessary for "success", the bound $F(\underline{x}_k)-F(\underline{x}_k+\underline{d}_k)>0.1\gamma\eta_{\check{\kappa}}$ is achieved. The argument of the previous paragraph completes the proof of Lemma 4 in this case, $\widehat{c}$ being the positive number $0.1\gamma\eta_{\check{\kappa}}$.

For the remainder of the proof, we restrict attention to the alternative case when every η_k is zero for the current ρ. Then, as stated in Section 2, the "trust region" steps are required to be "exact", and the algorithms set $Q_{k+1}\equiv Q_k$, which is reasonable when $F(\underline{x}_k+\underline{d}_k)=Q_k(\underline{x}_k+\underline{d}_k)$ occurs. Hence all iterations with the current ρ employ $Q_k\equiv Q_\kappa$, and each change $F(\underline{x}_k)-F(\underline{x}_k+\underline{d}_k)$ is the same as the predicted change $Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)$. Thus the conditions (2.4) and (2.7) for a "successful" trust region step reduce to the inequalities

$$Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)\ >\ 0 \qquad \text{and} \qquad \|\underline{d}_k\|\ \geq\ \tfrac{1}{2}\,\rho. \tag{6.11}$$

We introduce the notation $\lambda_1\leq\lambda_2\leq\ldots\leq\lambda_n$ for the eigenvalues of $G_k=\nabla^2 Q_\kappa$, arranged in ascending order. If λ_1 is negative, then an "exact" trust region step $\underline{d}_k$ satisfies $\|\underline{d}_k\|=\rho$ and $Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)\geq\frac{1}{2}\,|\lambda_1|\,\rho^2$. It follows from the first paragraph of this proof that the number of iterations with the current ρ is finite, $\widehat{c}$ being the positive number $\frac{1}{2}\,|\lambda_1|\,\rho^2$. Another possibility is that λ_1 is zero and $Q_\kappa(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is not bounded below. Then the gradient $\underline{g}_\kappa=\underline{\nabla}Q_\kappa(\underline{x}_\kappa)$ is not in the linear space spanned by the eigenvectors of $\nabla^2 Q_\kappa$ with nonzero eigenvalues. In other words, there is a vector, $\underline{w}$ say, in the null space of $\nabla^2 Q_\kappa$ that satisfies $\|\underline{w}\|=1$ and $\underline{w}^T\underline{g}_\kappa>0$. The null space property supplies the condition

$$\underline{w}^T\underline{g}_k\ =\ \underline{w}^T\left(\underline{g}_\kappa+\nabla^2 Q_\kappa\,(\underline{x}_k-\underline{x}_\kappa)\right)\ =\ \underline{w}^T\underline{g}_\kappa\ >\ 0 \tag{6.12}$$

for every relevant iteration number k. Furthermore, if $\underline{d}_k$ is an "exact" trust region step, then $Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)$ is at least $\underline{w}^T\underline{g}_\kappa\,\rho$, because Q_k is a linear polynomial along the direction $\underline{w}$. Again it follows from the first paragraph of this proof that Lemma 4 is true, $\widehat{c}$ being the number $\underline{w}^T\underline{g}_\kappa\,\rho$.

The only remaining situation is when the model $Q_k(\underline{x})=Q_\kappa(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is bounded below, but, as in the example when F is the function (6.10), some of the eigenvalues of $G_k=\nabla^2 Q_\kappa$ may be zero. If, during the iterations with the current ρ, a point $\underline{x}_k$ occurs such that $Q_\kappa(\underline{x}_k)$ is the least value of $Q_\kappa(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, then the first of the conditions (6.11) must fail, and formula (1.2) provides $\underline{x}_{k+1}=\underline{x}_k$. Thus, by induction, there are no more trust region steps with the current ρ, so termination occurs within the next $3n+3$ iterations. This argument covers the case when Q_κ is a constant function. For the remainder of the proof, we let λ_ℓ be the least positive eigenvalue of $\nabla^2 Q_\kappa$.

We consider an attempt at a "trust region" step when the gradient $\underline{g}_k=\underline{\nabla}Q_\kappa(\underline{x}_k)$ satisfies $\|\underline{g}_k\|\geq\rho\lambda_\ell$. The direction $\underline{d}=-\underline{g}_k/\|\underline{g}_k\|$ is the multiple of the Cauchy step that has length one, and the inequality

$$\begin{aligned}\tfrac{d}{d\theta}\,Q_\kappa(\underline{x}_k+\theta\,\underline{d})\ &=\ \underline{d}^T\underline{\nabla}Q_\kappa(\underline{x}_k+\theta\,\underline{d})\ =\ \underline{d}^T(\underline{g}_k+\theta\,\nabla^2 Q_\kappa\,\underline{d})\\ &\leq\ -\|\underline{g}_k\|+\theta\,\Omega\ \leq\ -\rho\,\lambda_\ell+\theta\,\Omega, \qquad \theta\in\mathcal{R},\end{aligned} \tag{6.13}$$

shows that the quadratic function $Q_\kappa(\underline{x}_k+\theta\underline{d})$, $\theta\in\mathcal{R}$, decreases monotonically when θ is in the interval $[0,\rho\lambda_\ell/\Omega]$, where Ω is still an upper bound on $\|\nabla^2 Q_\kappa\|$. We pick $\theta=\rho\lambda_\ell/\Omega$, which provides $\|\theta\underline{d}\|\leq\rho$, due to $\lambda_\ell\leq\Omega$ and $\|\underline{d}\|=1$. It follows that an "exact" trust region step $\underline{d}_k$ would achieve the condition

$$\begin{aligned} Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k) &\geq Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\theta\,\underline{d}) \geq -\tfrac{1}{2}\,\theta\,\underline{d}^T\underline{g}_k \\ &= \tfrac{1}{2}\,\|\underline{g}_k\|\,\rho\,\lambda_\ell\,/\,\Omega \geq \tfrac{1}{2}\,\rho^2\,\lambda_\ell^2\,/\,\Omega. \end{aligned} \tag{6.14}$$

Thus, by analogy with the argument in the first paragraph of this proof in the case $\widehat{c}=\frac{1}{2}\rho^2\lambda_\ell^2/\Omega$, the number of "successful" trust region steps $\underline{d}_k$ with $\|\underline{g}_k\|\geq\rho\lambda_\ell$ is finite.

This conclusion implies that, for all sufficiently large k with the current ρ, every "successful" trust region step $\underline{d}_k$ requires $\|\underline{g}_k\|<\rho\lambda_\ell$. Let $\underline{d}_k$ be "successful" in this setting, which happens during every run of $3n+3$ consecutive iterations if termination does not occur. We investigate vectors $\underline{d}$ that satisfy the equation

$$\underline{\nabla}Q_\kappa(\underline{x}_k+\underline{d}) \;=\; \underline{g}_k+\nabla^2 Q_\kappa\,\underline{d} \;=\; 0. \tag{6.15}$$

They exist because $\underline{g}_k$ is in the range space of $\nabla^2 Q_\kappa$ when $Q_\kappa(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is bounded below. Further, we make $\underline{d}$ unique when $\nabla^2 Q_\kappa$ is singular by requiring it to be in the range space of $\nabla^2 Q_\kappa$ too. Thus the identity (6.15) provides $\|\underline{g}_k\|=\|\nabla^2 Q_\kappa\,\underline{d}\|\geq\lambda_\ell\|\underline{d}\|$, so the condition $\|\underline{g}_k\|<\rho\lambda_\ell$ supplies $\|\underline{d}\|<\rho$. It follows from equation (6.15) that the "successful" step $\underline{d}_k$ can satisfy $\underline{\nabla}Q_\kappa(\underline{x}_k+\underline{d}_k)=0$. This actually happens, because $Q_\kappa(\underline{x}_k+\underline{d}_k)$ is the least value of the convex function $Q_\kappa(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, if and only if $\underline{\nabla}Q_\kappa(\underline{x}_k+\underline{d}_k)$ is zero. Then, because of remarks in the paragraph between expressions (6.12) and (6.13), no more "trust region" steps are possible with the current ρ. Therefore termination occurs after at most n more "alpha" steps and n more "beta" steps, with no further changes to the position of the trust region centre. The proof is complete. QED

7. Strong convergence

The main conclusion of the previous section is that, if the given conditions on the objective function hold, and if our algorithms take an infinite number of iterations, then the limit

$$\liminf\;\{\,\|\underline{\nabla}F(\underline{x}_k)\|\;:\;k=1,2,3,\ldots\,\}\;=\;0 \tag{7.1}$$

is achieved. It is proved below that "lim inf" can be replaced by "lim" without making any more assumptions.

Theorem 1 Let any algorithm from Section 2 be applied to a function F that, as usual, is bounded below and has bounded second derivatives, and let the number of iterations be infinite. Then, as $k\to\infty$, the gradients $\underline{\nabla}F(\underline{x}_k)$, $k=1,2,3,\ldots$, converge to the zero vector in $\mathcal{R}^n$.

Proof Let ε be any positive number. It is sufficient to prove that the condition $\|\underline{\nabla} F(\underline{x}_k)\| \leq 10\,\varepsilon$ is satisfied for all sufficiently large values of k. We let $k(\varepsilon)$ be the least positive integer such that the trust region radius ρ of the $k(\varepsilon)$-th iteration has the property

$$\rho \;\leq\; \tfrac{5}{9}\,\varepsilon \,/\, c_8, \tag{7.2}$$

where c_8 is the constant (6.1). We recall from Section 2 that the trust region radius is never increased, and that all reductions are by a factor of 10. It follows from Lemma 4 that $k(\varepsilon)$ is well defined. Further, $k \geq k(\varepsilon)$ is both necessary and sufficient for inequality (7.2) to hold on the k-th iteration.

Let the iteration number k satisfy the conditions $k \geq k(\varepsilon)$ and $\|\underline{\nabla} F(\underline{x}_k)\| \geq \varepsilon$. The property (7.2) gives $\|\underline{\nabla} F(\underline{x}_k)\| \geq \frac{9}{5} c_8 \rho > c_8 \rho$. Thus, as stated in the third paragraph of Section 6, if the k-th iteration tries to take a "trust region" step, then the attempt is "successful". In this case expression (6.7) is valid and it supplies the inequality

$$\begin{aligned} Q_k(\underline{x}_k) - Q_k(\underline{x}_k + \underline{d}_k) \;&\geq\; \rho\,\|\underline{\nabla} F(\underline{x}_k)\| - (c_3 + \tfrac{1}{2}\,\Omega)\,\rho^2 \\ &\geq\; \rho\,\varepsilon\left\{1 - \tfrac{5}{9}\,(c_3 + \tfrac{1}{2}\,\Omega) \,/\, c_8\right\}, \end{aligned} \tag{7.3}$$

the last line being derived from $\|\underline{\nabla} F(\underline{x}_k)\| \geq \varepsilon$ and $\rho \leq \frac{5}{9}\varepsilon / c_8$. Now definition (6.1) provides $(c_3 + \frac{1}{2}\Omega) \leq \frac{9}{19} c_8$, and the condition (2.4) is necessary for the "success" of a "trust region" step. Therefore the iteration under consideration achieves a decrease in the objective function that has the lower bound

$$F(\underline{x}_k) - F(\underline{x}_{k+1}) \;\geq\; 0.1\,\{Q_k(\underline{x}_k) - Q_k(\underline{x}_k + \underline{d}_k)\} \;\geq\; \tfrac{7}{95}\,\rho\,\varepsilon. \tag{7.4}$$

Another advantage of the property (7.2) is that it gives the condition

$$\|\underline{\nabla} F(\underline{x}_{k+1}) - \underline{\nabla} F(\underline{x}_k)\| \;\leq\; \varepsilon, \qquad k \geq k(\varepsilon). \tag{7.5}$$

Indeed, formula (1.2) with $\|\underline{d}_k\| \leq \rho$ imply $\|\underline{x}_{k+1} - \underline{x}_k\| \leq \rho$, so the upper bound Φ on $\|\nabla^2 F\|$ with inequality (7.2) yield the relation

$$\|\underline{\nabla} F(\underline{x}_{k+1}) - \underline{\nabla} F(\underline{x}_k)\| \;\leq\; \Phi\,\rho \;\leq\; \tfrac{5}{9}\,\varepsilon\,\Phi \,/\, c_8. \tag{7.6}$$

The assertion (7.5) follows from the remark that the definition (6.1) includes $c_8 \geq \frac{5}{9}\Phi$.

Let ℓ be any integer that satisfies $\ell \geq k(\varepsilon) + 2$ and $\|\underline{\nabla} F(\underline{x}_\ell)\| > 10\varepsilon$. We can assume that ℓ exists because otherwise there is nothing more to prove. Further, let m be the least integer that satisfies $m \geq \ell$ and $\|\underline{\nabla} F(\underline{x}_{m+1})\| < \varepsilon$, which also exists, because of the limit (7.1). We consider the iterations whose numbers k are in the interval $\ell \leq k \leq m$. Condition (7.5) implies that m is at least $\ell + 9$.

The following argument shows that, during these iterations, the trust region radius ρ remains constant. If this statement were false, then, because of the criterion for reducing ρ, given at the end of Section 2, the j-th iteration would have to make an "unsuccessful" attempt at a "trust region" step with the current

ρ for some integer j that satisfies $\max[\ell-2,\kappa]\leq j\leq m$, where κ is still the number of the first iteration that employs the current ρ. Our choices of ℓ and m, however, with the bound (7.5), provide $j\geq k(\varepsilon)$ and $\|\underline{\nabla}F(\underline{x}_j)\|\geq\varepsilon$ in all of these cases, so the possibility of an "unsuccessful" attempt at a "trust region" step has been excluded already in the second paragraph of this proof.

We require a lower bound on the number, ν say, of "trust region" steps $\underline{d}_k$ when k runs through the interval $[\ell,m]$. The total number of iterations for these values of k is $m-\ell+1$, and every 3 consecutive iterations include at least one attempt at a "trust region" step, all of them being "successful". Thus ν is at least $(m-\ell-1)/3$. The elementary relation

$$\begin{aligned}\|\underline{\nabla}F(\underline{x}_{m+1})-\underline{\nabla}F(\underline{x}_\ell)\| &\leq \textstyle\sum_{k=\ell}^{m}\|\underline{\nabla}F(\underline{x}_{k+1})-\underline{\nabla}F(\underline{x}_k)\| \\ &\leq (m-\ell+1)\,\Phi\,\rho \end{aligned} \tag{7.7}$$

gives a lower bound on $m-\ell$, the second line being taken from the first part of expression (7.6). Moreover, because ℓ and m satisfy $\|\underline{\nabla}F(\underline{x}_\ell)\|>10\varepsilon$ and $\|\underline{\nabla}F(\underline{x}_{m+1})\|<\varepsilon$, we have set up the inequality

$$\|\underline{\nabla}F(\underline{x}_{m+1})-\underline{\nabla}F(\underline{x}_\ell)\| > 9\,\varepsilon. \tag{7.8}$$

These remarks supply the condition

$$\nu \geq \frac{m-\ell-1}{3} = \left(1-\frac{2}{m-\ell+1}\right)\frac{m-\ell+1}{3} > \left(1-\frac{2}{m-\ell+1}\right)\frac{3\,\varepsilon}{\Phi\,\rho}. \tag{7.9}$$

Already we have noted that $m\geq\ell+9$ holds, which gives $m-\ell+1\geq 10$. It follows that ν is bounded below by the positive number $(12\,\varepsilon)/(5\,\Phi\rho)$.

The reduction (7.4) is achieved on each of the ν "trust region" iterations addressed in the previous paragraph. It follows from the lower bound on ν that the monotonically decreasing sequence $F(\underline{x}_k)$, $k=1,2,3\ldots$, has the property

$$F(\underline{x}_\ell)-F(\underline{x}_{m+1}) \geq (84\,\varepsilon^2)\,/\,(475\,\Phi). \tag{7.10}$$

We try to choose several $\{\ell,m\}$ pairs recursively, letting the first ℓ be the least integer that satisfies $\ell\geq k(\varepsilon)+2$ and $\|\underline{\nabla}F(\underline{x}_\ell)\|>10\,\varepsilon$, and letting each subsequent ℓ be the least integer that satisfies $\|\underline{\nabla}F(\underline{x}_\ell)\|>10\,\varepsilon$ and that is greater than the most recent previous value of m. Reductions in ρ are expected in some of the intervals between m and the next ℓ. The number of $\{\ell,m\}$ pairs generated by this recursion must be finite, because otherwise inequality (7.10) would contradict the assumption that F is bounded below. Therefore the condition $\|\underline{\nabla}F(\underline{x}_k)\|\leq 10\,\varepsilon$ holds for all sufficiently large k, which is the required result. QED

8. Quadratic models and numerical results

Some introductory numerical results are presented in this section that investigate gains in efficiency when linear models are replaced by quadratic ones in a version

of our algorithms. As stated in Section 2, this version employs the parameter settings $\alpha=0.1$, $\beta=5$, $\gamma=0.01$, $\tau_\alpha=1$ and $\tau_\beta=5$. These values were chosen before beginning the calculations reported below, so the parameters have not been tuned. Further, apart from minor changes of wording, the writing of Sections 1 to 7 was completed before starting work on the calculations.

The version sets $Q_{k+1}\equiv Q_k$ whenever possible, which happens if and only if $F(\underline{x}_k+\underline{d}_k)=Q_k(\underline{x}_k+\underline{d}_k)$ occurs. Otherwise, in the usual case $F(\underline{x}_k+\underline{d}_k)\neq Q_k(\underline{x}_k+\underline{d}_k)$, the equation

$$Q_{k+1}(\underline{x})-Q_k(\underline{x}) \;=\; \{F(\underline{x}_k+\underline{d}_k)-Q_k(\underline{x}_k+\underline{d}_k)\}\,\Lambda(\underline{x}), \qquad \underline{x}\in\mathcal{R}^n, \tag{8.1}$$

defines a function Λ that is a linear or quadratic polynomial. The interpolation conditions $Q_{k+1}(\underline{x}_k+\underline{d}_k)=F(\underline{x}_k+\underline{d}_k)$ and $Q_{k+1}(\underline{y}_i)=F(\underline{y}_i)$, $i\in\{0,1,\ldots,n\}\backslash\{t\}$, are satisfied. Therefore, because of the conditions (1.3) on Q_k, the function Λ has the properties

$$\Lambda(\underline{x}_k+\underline{d}_k) \;=\; 1 \qquad \text{and} \qquad \Lambda(\underline{y}_i) \;=\; 0, \quad i\in\{0,1,\ldots,n\}\backslash\{t\}. \tag{8.2}$$

We take the view that, if $F(\underline{x}_k+\underline{d}_k)\neq Q_k(\underline{x}_k+\underline{d}_k)$ holds, then Q_{k+1} is constructed in the following way. We pick a linear or quadratic polynomial Λ that obeys the Lagrange equations (8.2). Then we obtain Q_{k+1} by applying formula (8.1). The conditions (8.2) define Λ uniquely when it is a linear polynomial, because of the nonsingularity of the matrix (1.4) on the $(k+1)$-th iteration.

Also, when Λ is a linear polynomial, the value $\Lambda(\underline{y}_t)$ is nonzero. Indeed, if this assertion were false, then the conditions $\Lambda(\underline{y}_i)=0$, $i=0,1,\ldots,n$, with the nonsingularity of the matrix (1.4) on the k-th iteration would imply $\Lambda\equiv 0$, which would contradict $\Lambda(\underline{x}_k+\underline{d}_k)=1$. It follows from $Q_k(\underline{y}_t)=F(\underline{y}_t)$ that, if $F(\underline{x}_k+\underline{d}_k)-Q_k(\underline{x}_k+\underline{d}_k)$ is nonzero in expression (8.1), and if Q_{k+1} achieves all the conditions

$$Q_{k+1}(\underline{x}_k+\underline{d}_k) \;=\; F(\underline{x}_k+\underline{d}_k) \quad \text{and} \quad Q_{k+1}(\underline{y}_i) \;=\; F(\underline{y}_i), \quad i=0,1,\ldots,n, \tag{8.3}$$

then Λ must be quadratic instead of linear. Thus the updating of the model provides Q_{k+1} with some second derivative information, and Λ is required to have the properties

$$\Lambda(\underline{x}_k+\underline{d}_k) \;=\; 1 \qquad \text{and} \qquad \Lambda(\underline{y}_i) \;=\; 0, \quad i\in\{0,1,\ldots,n\}. \tag{8.4}$$

One way of satisfying them is to let Λ_0 be any linear polynomial that takes the values $\Lambda_0(\underline{y}_t)=0$ and $\Lambda_0(\underline{x}_k+\underline{d}_k)=1$, and to let the quadratic Λ be the product of Λ_0 with the linear function Λ that is defined by the equations (8.2).

In the numerical experiments of this section on quadratic models, we fix the freedom in Λ by applying the symmetric Broyden method, which works very well in the NEWUOA software (Powell, 2006). Specifically, Λ is the quadratic polynomial such that $\|\nabla^2\Lambda\|_F$ is least subject to the conditions (8.4), where the subscript F denotes the Frobenius norm of a matrix. Thus the sum of squares

of the elements of the second derivative matrix $\nabla^2\Lambda$ is minimized subject to the linear constraints (8.4). This quadratic programming problem defines Λ uniquely, and its solution requires only $\mathcal{O}(n^2)$ computer operations when the matrix (3.10) is available. We note also that this Λ is independent of the choice of t. Another remarkable property of the symmetric Broyden method is that, if F happens to be quadratic, then, because the calculation of Q_{k+1} is a least squares projection of Q_k into an affine linear set that includes F, the reduction

$$\|\nabla^2 Q_{k+1}-\nabla^2 F\|_F^2 \;=\; \|\nabla^2 Q_k-\nabla^2 F\|_F^2 \;-\; \|\nabla^2 Q_{k+1}-\nabla^2 Q_k\|_F^2 \qquad (8.5)$$

is achieved in the error of the approximation $\nabla^2 Q \approx \nabla^2 F$. These errors hardly ever become small in practice, however, but it is highly useful that, if equation (8.5) holds on every iteration, then $\|\nabla^2 Q_{k+1}-\nabla^2 Q_k\|$ tends to zero as $k\to\infty$.

The data for our numerical experiments are an initial vector of variables $\underline{x}_0$, the initial and final values of the trust region radius, set to 10^{-1} and 10^{-6}, respectively, a subroutine that supplies $F(\underline{x})$ for any $\underline{x}$ in $\mathcal{R}^n$, and a switch that is "off" or "on", where "off" causes every model Q_k to be a linear polynomial, and where "on" causes models to be updated by the symmetric Broyden method of the previous paragraph. The $n+1$ function values $F(\underline{x}_0)$ and $F(\underline{x}_0+10^{-1}\underline{e}_i)$, $i=1,2,\ldots,n$, are calculated before the first iteration, where $\underline{e}_i$ is the i-th coordinate direction in $\mathcal{R}^n$. The model Q_1 of the first iteration is always the linear polynomial that interpolates these values, the initial set $\{\underline{y}_i : i=0,1,\ldots,n\}$ being composed of the points $\underline{x}_0$ and $\underline{x}_0+10^{-1}\underline{e}_i$, $i=1,2,\ldots,n$, with $\underline{y}_0$ satisfying condition (2.1).

Two of the favourite objective functions of the author are employed, namely the "trigonometric sum of squares"

$$F(\underline{x}) \;=\; \sum_{i=1}^{2n}\Big\{c_i-\sum_{j=1}^{n}\left[\,S_{ij}\,\sin(x_j/\sigma_j)+C_{ij}\,\cos(x_j/\sigma_j)\,\right]\Big\}^2, \qquad \underline{x}\in\mathcal{R}^n, \qquad (8.6)$$

and the "chained Rosenbrock" function

$$F(\underline{x}) \;=\; \sum_{j=1}^{n-1}\{4\,(x_j-x_{j+1}^2)^2+(1-x_{j+1})^2\}, \qquad \underline{x}\in\mathcal{R}^n. \qquad (8.7)$$

In expression (8.6), each S_{ij} and C_{ij} is a fixed random integer from $[-100,100]$, each σ_j is a random constant from $[1,10]$, and each c_i is defined by $F(\underline{x}_*)=0$, after choosing $\underline{x}_*$ randomly from $[-\pi,\pi]^n$. Further, the starting vector $\underline{x}_0$ in the case (8.6) is picked by letting the weighted differences $[\underline{x}_0-\underline{x}_*]_j/\sigma_j$, $j=1,2,\ldots,n$, be independent random numbers from $[-\pi/10,\pi/10]$, where $[\underline{x}_0-\underline{x}_*]_j$ is the j-th component of $\underline{x}_0-\underline{x}_*$. In the case (8.7), the components of $\underline{x}_0$ are random numbers from the logarithmic distribution on $[0.5,2]$, and all the components of the optimal vector of variables $\underline{x}_*$ are one.

Different choices of the random numbers provide five test problems for each n in the cases (8.6) and (8.7). The values of $\#F$ and the final error $\|\underline{x}_f-\underline{x}_*\|_\infty$ were recorded whenever an algorithm was applied to a test problem, where $\#F$ is

n	Switch "off"	Switch "on"
20	18667 – 32022	2813 – 6559
40	28700 – 37674	6158 – 8875
80	63271 – 77076	14630 – 16619
160	159351 – 250010	29278 – 36067
320	426316 – 529585	63693 – 69215

Table 1: Range of values of $\#F$ for the problem (8.6)

n	Switch "off"	Switch "on"
20	12345 – 18431	1223 – 2115
40	10465 – 27292	2926 – 3793
80	38592 – 52637	5393 – 7036
160	49710 – 132376	13171 – 16510
320	113636 – 233876	31516 – 44620

Table 2: Range of values of $\#F$ for the problem (8.7)

the total number of calls of the subroutine that supplies $F(\underline{x})$ and where $\underline{x}_f$ is the vector of variables returned by the algorithm. Tables 1 to 4 show the ranges of $\#F$ and of $\|\underline{x}_f-\underline{x}_*\|_\infty$ over the five test problems for both objective functions, as indicated in the captions of the tables. The left and right halves of the tables were calculated by an algorithm that employs linear or quadratic models, respectively. The rows of the tables are distinguished by the number of variables n, which runs through the set $\{10\times 2^\ell : \ell=1,2,3,4,5\}$.

The results in the tables are a triumph for the use of quadratic models instead of linear ones, the values of $\#F$ and $\|\underline{x}_f-\underline{x}_*\|_\infty$ being reduced usually by more than a factor of five, although in both cases the number of interpolation points $\underline{y}_i$, $i=0,1,\ldots,n$, and $\underline{x}_k+\underline{d}_k$ is the same on each iteration, and the definitions

n	Switch "off"	Switch "on"
20	7.9×10^{-5} – 2.2×10^{-4}	7.6×10^{-6} – 1.6×10^{-5}
40	6.8×10^{-5} – 1.2×10^{-4}	9.8×10^{-6} – 1.3×10^{-5}
80	7.1×10^{-5} – 1.6×10^{-4}	8.4×10^{-6} – 2.1×10^{-5}
160	2.2×10^{-4} – 1.1×10^{-3}	1.1×10^{-5} – 1.2×10^{-5}
320	1.0×10^{-3} – 3.8×10^{-3}	8.1×10^{-6} – 1.4×10^{-5}

Table 3: Range of values of $\|\underline{x}_f-\underline{x}_*\|_\infty$ for the problem (8.6)

n	Switch "off"	Switch "on"
20	1.0×10^{-4} – 1.4×10^{-4}	5.7×10^{-6} – 1.1×10^{-5}
40	7.2×10^{-5} – 1.1×10^{-4}	1.9×10^{-6} – 6.8×10^{-6}
80	1.4×10^{-4} – 2.6×10^{-4}	9.5×10^{-7} – 1.0×10^{-5}
160	3.8×10^{-4} – 9.4×10^{-4}	4.1×10^{-6} – 1.7×10^{-5}
320	1.3×10^{-3} – 2.2×10^{-3}	7.2×10^{-6} – 1.8×10^{-5}

Table 4: Range of values of $\|\underline{x}_f-\underline{x}_*\|_\infty$ for the problem (8.7)

of the "alpha" and "beta" steps are also the same. The calculation of a "trust region" step when the switch is "on" is taken from NEWUOA (Powell, 2006), the method being a combination of conjugate gradients and two dimensional searches round the boundary of the trust region, which usually requires only $\mathcal{O}(n^2)$ computer operations to make $Q_k(\underline{x}_k+\underline{d}_k)$ close to its optimal value. The author had not expected the linear and quadratic models to perform so well when there are hundreds of variables. Perhaps the test functions (8.6) and (8.7) are too easy.

Another consideration is that much better efficiency may be achieved by avoiding long sequences of changes to the variables that are unnecessarily small. In many algorithms, ρ is doubled automatically if $\underline{d}_k$ is of length ρ and if it provides a sufficiently large decrease in the objective function, the condition

$$F(\underline{x}_k+\underline{d}_k)\ \le\ F(\underline{x}_k)-0.7\,[Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)] \tag{8.8}$$

being typical. The trust region radius is reduced later when $F(\underline{x})-F(\underline{x}_k+\underline{d}_k)$ compares unfavourably with $Q_k(\underline{x}_k)-Q_k(\underline{x}_k+\underline{d}_k)$. The author is going to run some experiments that take the adjustment of ρ from NEWUOA. Then the parameter ρ of the algorithms of Section 2 becomes a lower bound on the trust region radius, which is reduced by a factor of ten, as in Section 2, when the iterations with the current lower bound are complete.

Our work may be important to constrained optimization without derivatives, if the objective and constraint functions can be calculated outside the feasible region. One can employ the present number of interpolation points on each iteration, with the given "alpha" and "beta" steps to maintain inequalities (1.5) and (1.6), and each new value of F can be included in the updating of quadratic models by the symmetric Broyden method, as described earlier in this section. This approach seems to be straightforward when all the constraints are linear, because the contributions from the constraints are confined to the calculation of "trust region" steps, which is a quadratic programming problem. Furthermore, it is hoped that the small number of interpolation points on each iteration may provide some useful new algorithms for nonlinear constraints.

References

A.R. Conn, K. Scheinberg and L.N. Vicente (1997), "On the convergence of derivative-free methods for unconstrained optimization", in *Approximation Theory and Optimization*, eds. M.D. Buhmann and A. Iserles, Cambridge University Press (Cambridge), pp. 83–108.

A.R. Conn, K. Scheinberg and L.N. Vicente (2009a), "Global convergence of general derivative-free trust-region algorithms to first and second order critical points", *SIAM J. Optim.*, Vol. 20, pp. 387–415.

A.R. Conn, K. Scheinberg and L.N. Vicente (2009b), *Introduction to Derivative-Free Optimization*, SIAM Publications (Philadelphia).

M.J.D. Powell (1994), "A direct search optimization method that models the objective and constraint functions by linear interpolation", in *Advances in Optimization and Numerical Analysis*, eds. Susana Gomez and Jean-Pierre Hennart, Kluwer Academic (Dordrecht), pp. 51–67.

M.J.D. Powell (2006), "The NEWUOA software for unconstrained optimization without derivatives", in *Large-Scale Optimization*, editors G. Di Pillo and M. Roma, Springer (New York), pp. 255–297.