DAMTP 2000/NA10

On the Lagrange functions of quadratic models that are defined by interpolation[1]

M.J.D. Powell

Abstract: Quadratic models are of fundamental importance to the efficiency of many optimization algorithms when second derivatives of the objective function influence the required values of the variables. They may be constructed by interpolation to function values for suitable choices of the interpolation points. We consider the Lagrange functions of this technique, because they have some highly useful properties. In particular, they show whether a change to an interpolation point preserves nonsingularity of the interpolation equations, and they provide a bound on the error of the quadratic model. Further, they can be updated efficiently when an interpolation point is moved. These features are explained. Then it is shown that the error bound can control the adjustment of a trust region radius in a way that gives excellent convergence properties in an algorithm for unconstrained minimization calculations. Finally, a convenient procedure for generating the initial interpolation points is described.

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge,
Silver Street,
Cambridge CB3 9EW,
England.

September, 2000.

[1]Dedicated to Laurence C.W. Dixon in celebration of his 65th birthday.

1. Introduction

We let $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, be the objective function of an unconstrained minimization calculation. Many algorithms for minimizing F achieve a fast rate of convergence by having the property of being highly efficient if F happens to be a quadratic polynomial. Therefore adjustments to the variables may be made by working with a quadratic model

$$Q(\underline{x}) = c_Q + \underline{g}_Q^T(\underline{x}-\underline{x}_0) + \tfrac{1}{2}\,(\underline{x}-\underline{x}_0)^T G_Q\,(\underline{x}-\underline{x}_0), \qquad \underline{x}\in\mathcal{R}^n, \tag{1.1}$$

that is an approximation to $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$. Here $\underline{x}_0$ is a point of $\mathcal{R}^n$ that is fixed before Q is chosen, so the parameters of Q are the real number $c_Q\in\mathcal{R}$, the vector $\underline{g}_Q\in\mathcal{R}^n$ and the $n\times n$ symmetric matrix G_Q. One can delete $\underline{x}_0$ by letting it be the zero vector in $\mathcal{R}^n$, but we prefer not to do so, because the presence of $\underline{x}_0$ can reduce the effects of rounding errors on changes to $Q(\underline{x})$ that result from typical changes to the variables $\underline{x}$.

We suppose that the function value $F(\underline{x})$ can be calculated for any $\underline{x}$ in $\mathcal{R}^n$, but that no derivatives of the objective function are available. Therefore a possible way of choosing Q is to pick points $\underline{x}_i$, $i=1,2,\ldots,m$, in $\mathcal{R}^n$, such that the parameters of Q are defined uniquely by the interpolation conditions

$$Q(\underline{x}_i) = F(\underline{x}_i), \qquad i=1,2,\ldots,m. \tag{1.2}$$

By counting the number of independent parameters, we find that the value of m is $\frac{1}{2}(n+1)(n+2)$. Further, the need for the system (1.2) to have a unique solution imposes restrictions on the positions of the points $\underline{x}_i$, $i=1,2,\ldots,m$, that cause difficulties to some algorithms.

If these positions are suitable, then, for every integer j in $[1,m]$, a unique quadratic polynomial $\ell_j(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is defined by the equations

$$\ell_j(\underline{x}_i) = \delta_{ij}, \qquad i=1,2,\ldots,m, \tag{1.3}$$

where δ_{ij} is the Kronecker delta. The polynomials ℓ_j, $j=1,2,\ldots,m$, are called the Lagrange functions of the system (1.2). They have the property that the quadratic model that is given by the system can be written in the form

$$Q(\underline{x}) = \sum_{j=1}^{m} F(\underline{x}_j)\,\ell_j(\underline{x}), \qquad \underline{x}\in\mathcal{R}^n, \tag{1.4}$$

which is elementary and easy to verify. The main purpose of this paper is to draw attention to three more properties of the Lagrange functions that are important to an algorithm for unconstrained minimization without derivatives that is being developed by the author. It is expected that the properties will also be useful to some other new algorithms.

The employment of these properties requires the calculation of $\ell_j(\underline{x})$, $j=1,2,\ldots,m$, at a value of $\underline{x}$ that is not in the set $\{\underline{x}_i : i=1,2,\ldots,m\}$. Therefore, for every integer j in $[1,m]$, the real number $c_j \in \mathcal{R}$, the vector $\underline{g}_j \in \mathcal{R}^n$, and the $n\times n$ symmetric matrix G_j of the quadratic Lagrange function

$$\ell_j(\underline{x}) = c_j + \underline{g}_j^T(\underline{x}-\underline{x}_0) + \tfrac{1}{2}\,(\underline{x}-\underline{x}_0)^T G_j\,(\underline{x}-\underline{x}_0), \qquad \underline{x}\in\mathcal{R}^n, \tag{1.5}$$

are found explicitly. The total number of these coefficients is m^2, because only the lower triangular part of each G_j is stored. Their availability allows the coefficients of Q to be obtained from formula (1.4) in $\mathcal{O}(m^2)$ operations.

This amount of work is quite onerous, due to the value $m=\frac{1}{2}(n+1)(n+2)$. On the other hand, the solution of the system (1.2) by a direct method may take $\mathcal{O}(m^3)$ operations. Now each iteration of the new algorithm of the author changes only one of the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$, which suggests the possibility of calculating the coefficients of the new Lagrange functions from those of the old ones. We will find in Section 2 that this updating can be done in $\mathcal{O}(m^2)$ operations, and that the given procedure enjoys excellent stability. Thus the work per iteration of the new algorithm is $\mathcal{O}(n^4)$, which is optimal if all the coefficients of the Lagrange functions are stored, because a typical iteration alters the value of every coefficient. Section 2 also addresses the change that occurs to the determinant of the matrix of the $m\times m$ system (1.2) when one of the interpolation points is replaced, because it shows another advantage that comes from the use of Lagrange functions.

In Section 3 we study the accuracy of the approximation $Q(\underline{x})\approx F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, where Q is still the quadratic model that is defined by the interpolation conditions (1.2). An estimate of the accuracy in terms of the derivatives of F has to include third order derivatives, because the approximation is exact if F is a quadratic, but nonzero errors occur in the model if F is a cubic polynomial. Specifically, for general $\underline{x}\in\mathcal{R}^n$, we will establish the bound

$$|\,Q(\underline{x}) - F(\underline{x})\,| \;\le\; \tfrac{1}{6}M \sum_{j=1}^{m} |\ell_j(\underline{x})|\;\|\underline{x}-\underline{x}_j\|^3, \tag{1.6}$$

where M is a number that depends on third derivatives of the objective function. Further, all the contributions to the right hand side of expression (1.6) from $\underline{x}$ and from $\underline{x}_j$, $j=1,2,\ldots,m$, are confined to the terms that are under the summation sign, and the vector norm is Euclidean.

Fortunately, numerical experiments show that it is highly useful to assume that condition (1.6) is satisfied for some real number M, even if F does not have any third derivatives. Let the quadratic model and the Lagrange functions be available, and let $F(\underline{x})$ be calculated for some $\underline{x}$ in $\mathcal{R}^n$. We see that both sides of inequality (1.6) are zero if $\underline{x}$ is an interpolation point, but we exclude this case because then the value of $F(\underline{x})$ would be known already. Therefore none of the distances $\|\underline{x}-\underline{x}_j\|$, $j=1,2,\ldots,m$, is zero. It follows from the elementary equation

$$\sum_{j=1}^{m} \ell_j(\underline{x}) = 1, \qquad \underline{x}\in\mathcal{R}^n, \tag{1.7}$$

that the value of the sum in expression (1.6) is positive. Thus this expression and the value of $F(\underline{x})$ provide a lower bound on M. We recommend letting M be the greatest lower bound of this kind that has occurred so far during the iterations of an optimization calculation, if condition (1.6) is applied to estimate the accuracy of the current quadratic model at a new vector of variables $\underline{x}$. It is straightforward to choose M in this way for general objective functions.

Section 4 demonstrates the usefulness of this technique to the author's new algorithm for unconstrained optimization. It presents some features of the algorithm and a numerical example, which suggest that a superlinear rate of convergence may be achievable by trust region methods that employ quadratic models defined by interpolation. Finally, two other topics are addressed in Section 5. They are the initial choice of the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$, so that their Lagrange functions can be formed easily, and the reduction that can be made in the number of interpolation points when F has a sparse second derivative matrix.

The idea of forming quadratic models by interpolation for optimization without derivatives was proposed by Winfield (1973), but his algorithm does not employ Lagrange functions. Instead the present work began with some numerical testing of a method that uses linear polynomial models for constrained calculations (Powell, 1994). That testing gave favourable results when most of the freedom in the variables is taken up by active constraints, but it also exposes some severe inadequacies when second derivative terms are important. Therefore the problem of choosing and adjusting the positions of interpolation points for quadratic models was considered. It was noticed that the replacement of the interpolation point $\underline{x}_j$ by $\hat{\underline{x}}$, say, preserves nonsingularity of the system (1.2) if and only if $\ell_j(\hat{\underline{x}})$ is nonzero, where ℓ_j is still the Lagrange function that is defined by $\ell_j(\underline{x}_i)=\delta_{ij}$, $i=1,2,\ldots,m$. That remark was the basis of some new techniques that were described in a talk by the author at the 5th Stockholm Optimization Days in 1994, and that are mentioned in a survey paper (Powell, 1998). A few of those techniques influenced the development of the algorithm of Conn, Scheinberg and Toint (1997a, 1997b), which also employs quadratic models for optimization without derivatives, but that algorithm works with functions that are an alternative to Lagrange functions. The author, however, has always preferred Lagrange functions, because of the properties that are given in Section 2. Further, this point of view has been strengthened greatly by the importance of Lagrange functions to the numerical results that are reported in Section 4.

2. The determinant property and the updating of Lagrange functions

In order to study the system (1.2) of interpolation equations, we let b_j, $j=1,2,\ldots,m$, be any basis of the linear space of quadratic polynomials from $\mathcal{R}^n$ to $\mathcal{R}$. Therefore the current quadratic model can be written in the form

$$Q(\underline{x}) = \sum_{j=1}^{m} \mu_j \, b_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^n. \tag{2.1}$$

The coefficients μ_j, $j=1,2,\ldots,m$, are defined by the interpolation conditions. Specifically, they satisfy the equations

$$B\underline{\mu} = \underline{f}, \tag{2.2}$$

where B is the matrix that has the elements $B_{ij} = b_j(\underline{x}_i)$, $1 \le i,j \le m$, and where $\underline{\mu}$ and $\underline{f}$ are the vectors in $\mathcal{R}^m$ whose components are μ_j, $j=1,2,\ldots,m$, and $F(\underline{x}_i)$, $i=1,2,\ldots,m$, respectively. Of course the positions of the interpolation points have to be such that B is nonsingular. This condition is independent of the choice of basis, because it is equivalent to the requirement that the system (1.2) defines Q uniquely.

We assume that this condition holds, and that the optimization algorithm replaces one of the interpolation points. It is convenient to let $\underline{x}_1$ be replaced by $\underline{\hat{x}}$, say, and we ask whether the new matrix B is nonsingular. Therefore we introduce the notation $B(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$ to show the dependence of B on the positions of the interpolation points in $\mathcal{R}^n$, and we take the view that the ratio

$$\ell(\underline{\hat{x}}) = \det B(\underline{\hat{x}}, \underline{x}_2, \ldots, \underline{x}_m) \,/\, \det B(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m), \tag{2.3}$$

is a function of $\underline{\hat{x}} \in \mathcal{R}^n$. Now the denominator of expression (2.3) is independent of $\underline{\hat{x}}$, and the numerator is the determinant

$$\det \begin{pmatrix} b_1(\underline{\hat{x}}) & b_2(\underline{\hat{x}}) & \cdots & b_m(\underline{\hat{x}}) \\ b_1(\underline{x}_2) & b_2(\underline{x}_2) & \cdots & b_m(\underline{x}_2) \\ \vdots & \vdots & & \vdots \\ b_1(\underline{x}_m) & b_2(\underline{x}_m) & \cdots & b_m(\underline{x}_m) \end{pmatrix}, \qquad \underline{\hat{x}} \in \mathcal{R}^n. \tag{2.4}$$

It follows that $\ell(\underline{\hat{x}})$, $\underline{\hat{x}} \in \mathcal{R}^n$, is a quadratic polynomial that satisfies $\ell(\underline{x}_i) = 0$, $i=2,3,\ldots,m$. Moreover, the ratio (2.3) has the property $\ell(\underline{x}_1) = 1$. Therefore ℓ is the Lagrange function $\ell_1(\underline{x})$, $\underline{x} \in \mathcal{R}^n$.

Thus we deduce that, if $\underline{x}_j$ is replaced by $\underline{\hat{x}}$, where j is any integer from $[1,m]$, then the determinant of the matrix of the system (1.2) is multiplied by the factor $\ell_j(\underline{\hat{x}})$. The definition of Lagrange functions implies that this factor is independent of the basis $b_1, b_2, \ldots, b_m$. Furthermore, if the algorithm has calculated $F(\underline{\hat{x}})$ for some vector of variables $\underline{\hat{x}}$, and if the next quadratic model should satisfy

$Q(\underline{\hat{x}}) = F(\underline{\hat{x}})$, then we let $\underline{\hat{x}}$ become an interpolation point instead of $\underline{x}_j$, say. Because the choice of j should ensure that the new interpolation equations are well-conditioned, we require $|\ell_j(\underline{\hat{x}})|$ to be a relatively large member of the set $\{|\ell_i(\underline{\hat{x}})| \,:\, i=1,2,\ldots,m\}$. Therefore it is helpful if the Lagrange functions are available explicitly.

Next we address the updating of the Lagrange functions when the interpolation point $\underline{x}_j$ is replaced by $\underline{\hat{x}}$, say. Let the new Lagrange functions be $\hat{\ell}_i$, $i=1,2,\ldots,m$. Because $\hat{\ell}_j$ has to vanish at the interpolation points that are retained, it is a multiple of ℓ_j, and the condition $\hat{\ell}_j(\underline{\hat{x}})=1$ implies the normalization

$$\hat{\ell}_j(\underline{x}) = \ell_j(\underline{x}) \,/\, \ell_j(\underline{\hat{x}}), \qquad \underline{x} \in \mathcal{R}^n. \tag{2.5}$$

Moreover, because the values of ℓ_i and $\hat{\ell}_i$ are the same at all the interpolation points that are retained, where i is any integer from $[1,m]$, the difference $\hat{\ell}_i - \ell_i$ is a multiple of $\hat{\ell}_j$. Thus, for $i \neq j$, we deduce from $\hat{\ell}_i(\underline{\hat{x}})=0$ that $\hat{\ell}_i$ is the function

$$\hat{\ell}_i(\underline{x}) = \ell_i(\underline{x}) - \ell_i(\underline{\hat{x}})\, \hat{\ell}_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^n. \tag{2.6}$$

Equation (2.5) shows that the coefficients of $\hat{\ell}_j$ are those of ℓ_j divided by $\ell_j(\underline{\hat{x}})$, while, for $i \neq j$, equation (2.6) shows that the coefficients of $\hat{\ell}_i$ are those of ℓ_i minus $\ell_i(\underline{\hat{x}})$ times the corresponding coefficients of $\hat{\ell}_j$. The numbers $\ell_i(\underline{\hat{x}})$, $i=1,2,\ldots,m$, are usually available because of the choice of j. Therefore this updating procedure is easy to implement.

One way of investigating the stability of the procedure is to ask what happens if the functions ℓ_i, $i=1,2,\ldots,m$, are linearly independent quadratic polynomials that fail to satisfy the Lagrange conditions (1.3). In all cases the choice of j is required to have the property $\ell_j(\underline{\hat{x}}) \neq 0$, where j is still the index of the interpolation point that is replaced by $\underline{\hat{x}}$. Then formulae (2.5) and (2.6) are applied, and they give the conditions $\hat{\ell}_i(\underline{\hat{x}}) = \delta_{ij}$, $i=1,2,\ldots,m$, for general functions ℓ_i, $i=1,2,\ldots,m$. Similarly, if $\underline{x}_k$ is the new interpolation point of the previous iteration, then the use of formulae (2.5) and (2.6) on that iteration provided the values

$$\ell_i(\underline{x}_k) = \delta_{ik}, \qquad i=1,2,\ldots,m. \tag{2.7}$$

Therefore we can usually let k be an integer in $[1,m]$, different from j, that has the properties (2.7). Now, for $i=j$, equations (2.7) and (2.5) with $k \neq j$ yield $\hat{\ell}_j(\underline{x}_k) = 0$. Further, for $i \neq j$, this remark and equations (2.6) and (2.7) give $\hat{\ell}_i(\underline{x}_k) = \ell_i(\underline{x}_k) = \delta_{ik}$. Thus the Lagrange conditions $\hat{\ell}_i(\underline{x}_k) = \delta_{ik}$, $i=1,2,\ldots,m$, are inherited from the values (2.7). By using this argument recursively, we find that expression (2.7) is preserved by the updating of all future iterations. It follows that, if ℓ_i, $i=1,2,\ldots,m$, is any basis of the space of quadratic polynomials initially, then, except for the effects of computer rounding errors, the sequence of iterations will have constructed true Lagrange functions as soon as all the initial interpolation points have been replaced. Incorrect initial choices of ℓ_i, $i=1,2,\ldots,m$, are not recommended, however, but they happened occasionally during the development of the author's algorithm by mistake. Then the recovery process that has just been described worked very nicely in practice.

We are going to derive another form of this stability property from the arbitrary basis functions b_k, $k=1,2,\ldots,m$, that have been introduced already. For $i=1,2,\ldots,m$, the equations

$$\ell_i(\underline{x}) = \sum_{k=1}^{m} H_{ik}\, b_k(\underline{x}) \quad \text{and} \quad \hat{\ell}_i(\underline{x}) = \sum_{k=1}^{m} \hat{H}_{ik}\, b_k(\underline{x}), \qquad \underline{x}\in\mathcal{R}^n, \tag{2.8}$$

express ℓ_i and $\hat{\ell}_i$ in terms of the basis functions, so H and $\hat{H}$ are $m\times m$ matrices of coefficients. Further, we recall that B is the $m\times m$ matrix that has the elements $B_{ik}=b_k(\underline{x}_i)$, and we let $\hat{B}$ have the elements $\hat{B}_{ik}=b_k(\underline{\hat{x}}_i)$, where $\underline{\hat{x}}_i$, $i=1,2,\ldots,m$, are the interpolation points after the updating has been done, so they are the vectors $\underline{\hat{x}}_j=\underline{\hat{x}}$ and $\underline{\hat{x}}_i=\underline{x}_i$, $i\neq j$. Now we deduce from expression (2.8) that the functions ℓ_i, $i=1,2,\ldots,m$, satisfy the Lagrange conditions $\ell_i(\underline{x}_p)=\delta_{ip}$, $1\le i,p\le m$, if and only if the product HB^T is the $m\times m$ identity matrix I. Therefore, when H is a general nonsingular matrix that provides $\ell_j(\underline{\hat{x}})\neq 0$, the analysis of the previous paragraph suggests that the magnitude of $\hat{H}\hat{B}^T-I$ is no greater than the magnitude of HB^T-I in some sense. An explicit result of this kind is as follows.

Theorem 1: Let the given updating procedure replace $\underline{x}_j$ by $\underline{\hat{x}}$, and let the matrices H, $\hat{H}$, B and $\hat{B}$ be taken from the previous paragraph. Then the elements of the matrices (B^T-H^{-1}) and $(\hat{B}^T-\hat{H}^{-1})$ are the same, except that the j-th column of $(\hat{B}^T-\hat{H}^{-1})$ is zero.

Proof: The independence of the basis functions b_k, $k=1,2,\ldots,m$, with equations (2.5) and (2.8) imply the values

$$\hat{H}_{jk} = H_{jk} \,/\, \ell_j(\underline{\hat{x}}), \qquad k=1,2,\ldots,m. \tag{2.9}$$

Further, for $i\neq j$, equations (2.6), (2.8) and (2.9) give the formula

$$\hat{H}_{ik} = H_{ik} - \ell_i(\underline{\hat{x}})\, H_{jk} \,/\, \ell_j(\underline{\hat{x}}), \qquad k=1,2,\ldots,m. \tag{2.10}$$

Therefore, for $i=1,2,\ldots,m$, the i-th row of $\hat{H}$ is the i-th row of H minus a multiple of the j-th row of H, which a relation of the form

$$\hat{H} = \left(I - \underline{v}\,\underline{e}_j^T\right) H, \tag{2.11}$$

for some $\underline{v}$ in $\mathcal{R}^m$, where $\underline{e}_j$ is the j-th coordinate vector in $\mathcal{R}^m$. It follows that $\hat{H}^{-1}$ is the matrix

$$\hat{H}^{-1} = H^{-1}\left(I + [\,1-\underline{e}_j^T\underline{v}\,]^{-1}\,\underline{v}\,\underline{e}_j^T\right), \tag{2.12}$$

which is well-defined, because the j-th component of $\underline{v}$ is $1-1/\ell_j(\underline{\hat{x}})$, so $[1-\underline{e}_j^T\underline{v}]^{-1}$ has the value $\ell_j(\underline{\hat{x}})$. Expression (2.12) provides the identities $\hat{H}^{-1}\underline{e}_k = H^{-1}\underline{e}_k$, $k\neq j$. In other words, any differences between $\hat{H}^{-1}$ and H^{-1} occur only in the j-th columns of these matrices. Moreover, the k-th columns of B^T and $\hat{B}^T$ are the same for $k\neq j$, because they have the elements $b_i(\underline{x}_k)$, $i=1,2,\ldots,m$, and $b_i(\underline{\hat{x}}_k)$, $i=1,2,\ldots,m$, respectively, and we recall $\underline{\hat{x}}_k=\underline{x}_k$, $k\neq j$. Therefore it remains to prove that the j-th column of $(\hat{B}^T-\hat{H}^{-1})$ is zero.

Using the definitions of $\hat{H}$ and $\hat{B}$, we deduce the values

$$\hat{\ell}_i(\underline{\hat{x}}) = \sum_{k=1}^m \hat{H}_{ik}\, b_k(\underline{\hat{x}}) = \sum_{k=1}^m \hat{H}_{ik}\, \hat{B}_{jk}, \qquad i=1,2,\ldots,m. \tag{2.13}$$

Thus the j-th column of $\hat{H}\hat{B}^T$ has the components $\hat{\ell}_i(\underline{\hat{x}})$, $i=1,2,\ldots,m$. Moreover, we have noted already that the updating procedure gives the Lagrange conditions $\hat{\ell}_i(\underline{\hat{x}})=\delta_{ij}$, $i=1,2,\ldots,m$. It follows that the j-th column of $\hat{H}\hat{B}^T$ is the j-th coordinate vector. In other words, the equation $(\hat{H}\hat{B}^T-I)\underline{e}_j=0$ is achieved. Multiplication of this equation by $\hat{H}^{-1}$ yields the identity $(\hat{B}^T-\hat{H}^{-1})\underline{e}_j=0$, which completes the proof. □

This theorem provides a property that is deduced in the paragraph that includes equation (2.7). Specifically, we consider the matrices (B^T-H^{-1}) of successive iterations when ℓ_i, $i=1,2,\ldots,m$, is any basis of the space of quadratic polynomials initially. The theorem states that the only changes to the matrices are that columns become zero, the indices of those columns being the indices of the interpolation points that are dropped. Therefore, when all of the initial interpolation points have been replaced, the equation $B^T=H^{-1}$ is satisfied, which means that the Lagrange conditions $HB^T=I$ are achieved. An advantage of this argument over the previous analysis is that it suggests that the conditions are obtained in a robust way. Indeed, a corollary of the theorem is that the sequence of Frobenius norms $\|B^T-H^{-1}\|_F$ decreases monotonically, even if the denominator $\ell_j(\underline{\hat{x}})$ of formula (2.5) is small. Numerical experiments have confirmed that the given updating procedure is highly successful at maintaining good accuracy.

3. An error bound for the quadratic model

We assume in this section that the objective function $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, has third derivatives that are bounded and continuous. Therefore, if $\underline{y}$ is any point in $\mathcal{R}^n$, and if $\underline{d}$ is any vector in $\mathcal{R}^n$ that has unit Euclidean length, which is the condition $\|\underline{d}\|=1$, then the function of one variable

$$\phi(\alpha) = F(\underline{y}+\alpha\,\underline{d}), \qquad \alpha\in\mathcal{R}, \tag{3.1}$$

also has bounded and continuous third derivatives. Further, there is a least nonnegative number M, independent of $\underline{y}$ and $\underline{d}$, such that every function of this form has the property

$$|\phi'''(\alpha)| \leq M, \qquad \alpha \in \mathcal{R}. \tag{3.2}$$

This value of M is suitable for the error bound (1.6), which is stated and proved as a theorem.

Theorem 2: Let the statements of the previous paragraph hold, and let the interpolation points $\underline{x}_i \in \mathcal{R}^n$, $i=1,2,\ldots,m$, be in any positions such that the equations (1.2) define a unique quadratic polynomial Q from $\mathcal{R}^n$ to $\mathcal{R}$. Then, for every $\underline{x}$ in $\mathcal{R}^n$, the error of the quadratic model satisfies the condition

$$|\, Q(\underline{x}) - F(\underline{x}) \,| \;\leq\; \tfrac{1}{6} M \sum_{j=1}^{m} |\ell_j(\underline{x})| \, \|\underline{x} - \underline{x}_j\|^3, \tag{3.3}$$

where the functions ℓ_j, $j=1,2,\ldots,m$, are the Lagrange functions of the interpolation equations.

Proof: We make any choice $\underline{y} \in \mathcal{R}^n$ of the vector of variables, we regard $\underline{y}$ as fixed for the moment, and we derive a bound on $|Q(\underline{y}) - F(\underline{y})|$. The Taylor series expansion of $F(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, about the point $\underline{y}$ is important. Specifically, we let $T(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, be the quadratic polynomial that contains all the zero order, first order and second order terms of this expansion, and we consider the possibility of replacing the objective function F by F–T. The replacement would preserve all the third derivatives of the objective function, because T is a quadratic polynomial. Therefore the given choice of M would remain valid. Furthermore, the quadratic model of $F-T$ that is defined by the interpolation method would be $Q-T$. It follows that the error of the new quadratic model of the new objective function is $F-Q$ as before. Therefore, when seeking a bound on $|Q(\underline{y}) - F(\underline{y})|$ in terms of third derivatives of the objective function, we assume without loss of generality that the function value $F(\underline{y})$, the gradient vector $\underline{\nabla} F(\underline{y})$ and the second derivative matrix $\nabla^2 F(\underline{y})$ are all zero.

Let j be any integer from $[1,m]$ such that $\underline{x}_j \neq \underline{y}$, let $\underline{d}$ be the vector

$$\underline{d} = (\underline{x}_j - \underline{y}) \,/\, \|\underline{y} - \underline{x}_j\|, \tag{3.4}$$

which has unit Euclidean length, and let $\phi(\alpha)$, $\alpha \in \mathcal{R}$, be the function (3.1). The Taylor series with explicit remainder gives the formula

$$\phi(\alpha) \;=\; \phi(0) + \alpha\,\phi'(0) + \tfrac{1}{2}\alpha^2 \phi''(0) + \tfrac{1}{6}\alpha^3 \phi'''(\xi), \qquad \alpha \geq 0, \tag{3.5}$$

where ξ depends on α and is in the interval $[0,\alpha]$. The values of $\phi(0)$, $\phi'(0)$ and $\phi''(0)$ are all zero due to the assumptions of the previous paragraph, and we pick $\alpha = \|\underline{y} - \underline{x}_j\|$. Thus expressions (3.4), (3.1), (3.5) and (3.2) provide the bound

$$|F(\underline{x}_j)| \;=\; \tfrac{1}{6}\alpha^3 \, |\phi'''(\xi)| \;\leq\; \tfrac{1}{6} M \, \|\underline{y} - \underline{x}_j\|^3, \tag{3.6}$$

which also holds without the middle part in the case $\underline{x}_j = \underline{y}$, because of the assumption $F(\underline{y})=0$. Using $F(\underline{y})=0$ again, we deduce from formula (1.4) and from inequality (3.6) that the error $Q(\underline{y})-F(\underline{y})$ has the property

$$\begin{aligned} |\, Q(\underline{y}) - F(\underline{y}) \,| \;=\; |Q(\underline{y})| \;&=\; |\, \sum_{j=1}^{m} F(\underline{x}_j)\, \ell_j(\underline{y}) \,| \\ &\leq\; \tfrac{1}{6} M \sum_{j=1}^{m} |\ell_j(\underline{y})| \, \|\underline{y}-\underline{x}_j\|^3. \end{aligned} \tag{3.7}$$

Therefore, because $\underline{y}$ is arbitrary, the theorem is true. □

The author discovered the error bound (3.3) himself, before enquiring whether it had been published already. The answer to that question was obtained with the help of Carl de Boor and Shayne Waldron, whose assistance is acknowledged gratefully. Specifically, a one variable $(n=1)$ version of Theorem 2 is stated in equation (3.7.10) of Davis (1975), and is called "Kowalewski's exact remainder". That result is generalized by Waldron (1998), who also gives references to related work.

4. A trust region technique for noisy functions

One expects objective functions to be unsmooth in practice because of computer rounding errors, but much larger anomalies may occur. For example, switching from one approximation to another by adaptive procedures in the calculation of $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, causes discontinuities, or there may be substantial contributions to $F(\underline{x})$ from random errors in some preliminary measurements, or there may be loss of accuracy because $F(\underline{x})$ depends on auxiliary functions whose values are confined to discrete sets. In all of these cases there are advantages in defining quadratic models by data fitting methods, but we are addressing the use of interpolation. Now the effects of errors in the function values $F(\underline{x}_i)$, $i=1,2,\ldots,m$, on the derivatives of our quadratic model become large when the interpolation points are close together. Therefore the algorithm of the author that has been mentioned is designed to reduce the objective function as much as possible using interpolation points that are well separated, before closer interpolation points are allowed.

The implementation of this strategy is governed by a trust region radius Δ, whose value gives the magnitude of least distances between interpolation points. Indeed, a new vector of variables, $\underline{\hat{x}}$ say, is usually the $\underline{x}$ that solves the problem

$$\text{minimize} \quad Q(\underline{x}) \quad \text{subject to} \quad \|\underline{x}-\underline{x}_k\| \leq \rho, \tag{4.1}$$

where ρ is a second trust region radius that satisfies $\rho \geq \Delta$, and where k is the integer in $[1,m]$ such that $F(\underline{x}_k)$ is the least of the function values $F(\underline{x}_i)$, $i=1,2,\ldots,m$, any ties being broken by preferring the $F(\underline{x}_k)$ that was calculated first. The view is taken, however, that $\underline{\hat{x}}$ is too close to $\underline{x}_k$ if $\|\underline{\hat{x}}-\underline{x}_k\|$ is less than $\frac{1}{2}\Delta$. In this case, and also if the previous iteration did not achieve a sufficient reduction in the objective function, there are several tests on the goodness of the current quadratic model. If any of them fail, then one interpolation point is moved to a new position that should improve the model. Otherwise, Δ is reduced, or termination occurs if Δ is at its final value. For example, Δ may change from $\Delta=10^{-1}$ to $\Delta=10^{-8}$ during the calculation, each reduction being by a factor of ten.

This kind of method is considered by Powell (1998), and there the tests on the goodness of the approximation $Q\approx F$ include the conditions

$$\|\underline{x}_j-\underline{x}_k\| \le \beta\Delta, \qquad j=1,2,\ldots,m, \tag{4.2}$$

where k has the value that is given in the previous paragraph, and where β is a constant that exceeds one, the choice $\beta=2$ being typical. It follows that, if Δ is reduced by a factor of ten, and if $\|\underline{x}_i-\underline{x}_j\|>0.1\beta\Delta$ occurs for every $i\neq j$ for the old value of Δ, then at least $m-1$ changes have to be made to the positions of the interpolation points in order to satisfy the tests (4.2) for the new Δ. Therefore the need to generate about $\frac{1}{2}n^2$ new values of the objective function was accepted. Then, however, the author did not know about our Theorem 2. We will find that it provides a technique that allows far fewer than $\frac{1}{2}n^2$ new values of F to be sufficient for each reduction in Δ, towards the end of some trial calculations, which were performed by the current version of the author's new algorithm.

The technique is based on the remark that there is no need to improve the quadratic model if the right hand side of the error bound (1.6) is "sufficiently small". In order to be specific, we recall that, if the solution $\underline{\hat{x}}$ of the problem (4.1) satisfies $\|\underline{\hat{x}}-\underline{x}_k\|<\frac{1}{2}\Delta$, then $\underline{\hat{x}}$ is assumed to be too close to $\underline{x}_k$. In this case $\underline{\hat{x}}$ must be the vector of variables that minimizes $Q(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, so the second derivative matrix of Q, namely G_Q, is positive definite (or semi-definite), and we are abandoning a possible reduction in the objective function that has the predicted value $\frac{1}{2}(\underline{x}_k-\underline{\hat{x}})^T G_Q(\underline{x}_k-\underline{\hat{x}})$. Therefore this expression provides a suitable tolerance on the errors of the model. We prefer, however, to be cautious about relaxing the conditions (4.2) if G_Q is nearly singular. Therefore we let the magnitude of "sufficiently small" in the tests of the error bound be the quantity

$$\varepsilon = \tfrac{1}{2}\Delta^2\max[0,\lambda_Q], \tag{4.3}$$

where λ_Q is the least eigenvalue of G_Q.

The value of M is chosen by the method of the paragraph that includes equation (1.7), and also some other features are included in the author's algorithm, because the work of testing the inequality

$$\tfrac{1}{6}M\sum_{j=1}^{m}|\ell_j(\underline{x})|\,\|\underline{x}-\underline{x}_j\|^3 \le \varepsilon \tag{4.4}$$

for a range of values of $\underline{x}$ would be prohibitive. One complication is that, if this left hand side is regarded as a function of $\underline{x}$, then there are discontinuities in first derivatives wherever any of the values $\ell_j(\underline{x})$, $j=1,2,\ldots,m$, is zero. Therefore each of the terms $\frac{1}{6}M\,|\ell_j(\underline{x})|\,\|\underline{x}-\underline{x}_j\|^3$, $j=1,2,\ldots,m$, is compared separately with expression (4.3), which gives an optimistic error bound that seems to be successful in practice. Further, because the algorithm gives particular attention to vectors $\underline{x}$ that satisfy $\|\underline{x}-\underline{x}_k\|\le\Delta$, it uses an estimate of the number $\max\{|\ell_j(\underline{x})| : \|\underline{x}-\underline{x}_k\|\le\Delta\}$ that we call θ_j, and it makes the approximation $\|\underline{x}-\underline{x}_j\|^3\approx\|\underline{x}_k-\underline{x}_j\|^3$. Another feature is that the tests (4.2) are also employed, which is important in the case $\varepsilon=0$. These ingredients provide the following criterion. It is assumed that the interpolation points provide an acceptably accurate quadratic model if at least one of the two conditions

$$\tfrac{1}{6}M\theta_j\,\|\underline{x}_j-\underline{x}_k\|^3\le\varepsilon \quad\text{and}\quad \|\underline{x}_j-\underline{x}_k\|\le\beta\Delta \tag{4.5}$$

is achieved for every integer j in $[1,m]$. Otherwise, if an improvement to the model is required, then an interpolation point $\underline{x}_j$ that fails both of the tests (4.5) is moved to a new position.

Powell (1998) recommends even more tests for the goodness of the quadratic model, including an assessment of the conditioning of the system (1.2) that may require the replacement of an interpolation point $\underline{x}_j$ that satisfies $\|\underline{x}_j-\underline{x}_k\|\le\beta\Delta$. All of those tests have been abandoned, however, because they increase the amount of computation substantially, and numerical results have indicated that they are unnecessary. It seems, therefore, that the current rules for choosing the positions of new interpolation points, and for deciding which one to drop if the solution $\hat{\underline{x}}$ of the problem (4.1) is to become an interpolation point, are very suitable. These and other details will be published by the author when the development of the new algorithm is complete.

The reason for considering a provisional version of the algorithm is that it gave the numerical results that were presented by the author at the 17th Symposium of the Mathematical Programming Society in Atlanta. In most of those calculations the objective function has the trigonometric form

$$F(\underline{x})=\sum_{i=1}^{n}\Big(b_i-\sum_{j=1}^{n}(S_{ij}\sin x_j+C_{ij}\cos x_j)\Big)^2,\qquad \underline{x}\in\mathcal{R}^n, \tag{4.6}$$

that was used originally by Fletcher and Powell (1963). Here the elements of the matrices S and C are random integers from $[-100,100]$, a vector $\underline{x}_*$ is generated randomly in order to define the numbers b_i, $i=1,2,\ldots,n$, by $F(\underline{x}_*)=0$, and the initial $\underline{x}$ is formed by adding random contributions from $[-0.1\pi,0.1\pi]$ to the components of $\underline{x}_*$. Further, as suggested in the second paragraph of this section, the trust region radius Δ is reduced from 10^{-1} to 10^{-8} during each calculation, each reduction being by a factor of ten. The value of n has to be small, because

$n=5$		Δ	$n=20$	
$\#F$	$F_{\min}$		$\#F$	$F_{\min}$
47	3.4×10^{-1}	10^{-1}	522	2.1×10^{0}
69	8.1×10^{-4}	10^{-2}	1210	4.3×10^{-4}
94	3.6×10^{-8}	10^{-3}	1264	4.2×10^{-6}
97	3.4×10^{-9}	10^{-4}	1431	9.4×10^{-10}
104	8.7×10^{-12}	10^{-5}	1448	2.7×10^{-11}
108	2.4×10^{-13}	10^{-6}	1462	2.6×10^{-13}
113	3.0×10^{-17}	10^{-7}	1475	4.3×10^{-15}
117	1.7×10^{-18}	10^{-8}	1489	3.3×10^{-18}

Table 1: Results for the objective function (4.6)

of the $\mathcal{O}(n^4)$ complexity of each iteration. Specifically, five different objective functions of the form (4.6) were generated for each of the choices $n=3$, $n=5$, $n=10$ and $n=20$.

Table 1 gives some results for the $n=5$ and $n=20$ calculations that required most values of the objective function. Each row of the table states the total number of function values so far, namely $\#F$, and the least calculated value of the objective function so far, namely $F_{\min}$, when the work for the current Δ is complete. The $F_{\min}$ columns show that excellent accuracy is achieved. We see also that the number of calculations of F that are made for each new value of Δ becomes much less than $\frac{1}{2}n^2$. Therefore a huge improvement has been gained by taking advantage of the bound (1.6) on the error of the quadratic model.

Tiny values of Δ that are less than the ones in Table 1, however, usually destroy these gains, because of a tendency for the errors in the calculation of $F(\underline{x})$ to become much larger than the true values of $|Q(\underline{x})-F(\underline{x})|$ that are relevant. Thus the automatic choice of M, which is derived from inequality (1.6), becomes huge. It follows that the first of the conditions (4.5) is hardly ever satisfied, which reintroduces the disadvantages of requiring all the tests (4.2) to hold before each reduction in Δ.

5. Initialization and sparsity

We are going to study the construction of the first quadratic model and its Lagrange functions, when an initial vector of variables, $\underline{x}_1 \in \mathcal{R}^n$ say, and the initial value of the trust region radius Δ are given. We let $\underline{x}_1$ be the first interpolation point, but suitable points $\underline{x}_i$, $i=2,3,\ldots,m$, for the equations (1.2) are to be chosen automatically. For each integer k in $[2, m]$, the position of $\underline{x}_k$ can depend not only on $\underline{x}_i$, $i = 1, 2, \ldots, k-1$, but also on the function values $F(\underline{x}_i)$, $i=1, 2, \ldots, k-1$. Indeed, these function values may provide an approximation to the objective function, for example a linear polynomial that is defined by interpolation when $k = n+2$. Then some of the freedom in $\underline{x}_k$ may be taken up by seeking a low value of the approximation at the new interpolation point, subject to a trust region bound. An interesting procedure of this kind is proposed by Conn and Toint (1996).

Another consideration is that we require all the coefficients of all the Lagrange functions of the initial interpolation points $\underline{x}_i$, $i=1, 2, \ldots, m$. This task is easy to perform in the new algorithm of the author, because the points are chosen so that most of the coefficients are zero. In order to describe the method that is used, we address the calculation of Q from the equations (1.2), before we expose the coefficients of the Lagrange functions.

Let $\underline{x}_0$ in expressions (1.1) and (1.5) be the given vector $\underline{x}_1$, and let $(\underline{g}_Q)_j$ and $(G_Q)_{jk}$ denote the j-th component of $\underline{g}_Q$ and the (j, k)-th element of the symmetric matrix G_Q, respectively. Then the parameter c_Q of expression (1.1) has the value $F(\underline{x}_1)$, and, for $j=1, 2, \ldots, n$, the parameters $(\underline{g}_Q)_j$ and $(G_Q)_{jj}$ are defined by the conditions $Q(\underline{x}_1)=F(\underline{x}_1)$, $Q(\underline{x}_{2j})=F(\underline{x}_{2j})$ and $Q(\underline{x}_{2j+1})=F(\underline{x}_{2j+1})$, if we pick the forms

$$\underline{x}_{2j} = \underline{x}_0 + \alpha_j\, \underline{e}_j \quad \text{and} \quad \underline{x}_{2j+1} = \underline{x}_0 + \beta_j\, \underline{e}_j, \qquad j=1, 2, \ldots, n, \tag{5.1}$$

where $\underline{e}_j$ is the j-th coordinate vector, and where α_j and β_j are different nonzero real numbers. For example, the algorithm of the author sets $\alpha_j = \Delta$ and $\beta_j = 2\Delta$ or $-\Delta$ for each j, the value $\beta_j = 2\Delta$ being preferred if $F(\underline{x}_{2j}) < F(\underline{x}_0)$ occurs. Furthermore, if $\underline{x}$ has the form $\underline{x}_0 + \gamma_j \underline{e}_j + \gamma_k \underline{e}_k$, $j \neq k$, then equation (1.1) provides the identity

$$Q(\underline{x}) = c_Q + \gamma_j\, (\underline{g}_Q)_j + \gamma_k\, (\underline{g}_Q)_k + \tfrac{1}{2}\gamma_j^2\, (G_Q)_{jj} + \gamma_j \gamma_k\, (G_Q)_{jk} + \tfrac{1}{2}\gamma_k^2\, (G_Q)_{kk}. \tag{5.2}$$

It follows from the choices made already that the only unknown parameter on the right hand side is $(G_Q)_{jk}$. Therefore its value is defined by the conditions (1.2), if we include the vector $\underline{x}_0 + \gamma_j \underline{e}_j + \gamma_k \underline{e}_k$ in the set $\{\underline{x}_i : i = 1, 2, \ldots, m\}$, and if both γ_j and γ_k are nonzero. We do so for all pairs of integers (j, k) that satisfy $1 \leq j < k \leq n$, which gives the interpolation points

$$\underline{x}_{i(j,k)} \;=\; \underline{x}_0 + \gamma_j\, \underline{e}_j + \gamma_k\, \underline{e}_k, \qquad 1 \leq j < k \leq n, \tag{5.3}$$

where $i(j, k) = 2n+1+j+\frac{1}{2}(k-1)(k-2)$. Thus the indices of the points (5.3) are all the integers in $[2n+2, m]$. The algorithm of the author sets $\gamma_j = \pm\Delta$, $j=1, 2, \ldots, n$, the plus sign being chosen in the case $F(\underline{x}_{2j}) < F(\underline{x}_0)$.

It is straightforward to identify the Lagrange functions of the interpolation points (5.1) and (5.3). Most of them are given by the remark that, if j and k

are any integers that satisfy $1 \leq j < k \leq n$, then, for $i = i(j,k)$, the quadratic polynomial

$$\ell_i(\underline{x}) = (\underline{x} - \underline{x}_0)_j \, (\underline{x} - \underline{x}_0)_k \,/\, (\gamma_j \gamma_k), \qquad \underline{x} \in \mathcal{R}^n, \tag{5.4}$$

enjoys the properties $\ell_i(\underline{x}_t) = \delta_{ti}$, $t = 1,2,\ldots,m$, where $(\underline{x} - \underline{x}_0)_j$ and $(\underline{x} - \underline{x}_0)_k$ are the j-th and k-th components of $\underline{x} - \underline{x}_0$. This construction provides the last $\frac{1}{2}n(n-1)$ Lagrange functions, and, because only the lower triangular parts of symmetric matrices are stored, each of them has only one nonzero parameter. These functions and the quadratic polynomials

$$\left.\begin{array}{l} \hat{\ell}_{2j}(\underline{x}) = (\underline{x} - \underline{x}_0)_j \, (\underline{x} - \underline{x}_{2j+1})_j \,/\, (\alpha_j^2 - \alpha_j \beta_j), \;\; \underline{x} \in \mathcal{R}^n \\ \hat{\ell}_{2j+1}(\underline{x}) = (\underline{x} - \underline{x}_0)_j \, (\underline{x} - \underline{x}_{2j})_j \,/\, (\beta_j^2 - \alpha_j \beta_j), \;\; \underline{x} \in \mathcal{R}^n \end{array}\right\}, \qquad j = 1,2,\ldots,n, \tag{5.5}$$

produce a useful formula for ℓ_i, $i = 2,3,\ldots,2n+1$. Indeed, because equation (5.1) implies that the polynomials (5.5) satisfy $\hat{\ell}_i(\underline{x}_t) = \delta_{ti}$, $t = 1,2,\ldots,2n+1$, the Lagrange conditions $\ell_i(\underline{x}_t) = \delta_{ti}$, $t = 1,2,\ldots,m$, are achieved by the functions

$$\ell_i(\underline{x}) = \hat{\ell}_i(\underline{x}) - \sum_{t=2n+2}^{m} \hat{\ell}_i(\underline{x}_t) \, \ell_t(\underline{x}), \qquad \underline{x} \in \mathcal{R}^n, \quad i = 2,3,\ldots,2n+1. \tag{5.6}$$

Then ℓ_1 can be written in the form

$$\ell_1(\underline{x}) = 1 - \sum_{t=2}^{m} \ell_t(\underline{x}), \qquad \underline{x} \in \mathcal{R}^n, \tag{5.7}$$

which completes the definitions of the Lagrange functions.

Equation (5.6) is useful, because the expansion of $\hat{\ell}_i(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, about $\underline{x}_0$ contains only one or two nonzero parameters, and because ℓ_t, $2n+2 \leq t \leq m$, has only one nonzero parameter. Further, letting j be the integer such that i is $2j$ or $2j+1$, we deduce from the definition (5.5) that $\hat{\ell}_i(\underline{x}_t)$ is nonzero only if the j-th component of $\underline{x}_t - \underline{x}_0$ is nonzero, which happens for just $n-1$ of the points (5.3). It follows that each of the functions (5.6) has at most $n+1$ nonzero parameters, but all of the parameters of ℓ_1 may be nonzero. These remarks show that all the nonzero coefficients of all the initial Lagrange functions can be generated in only $\mathcal{O}(n^2)$ computer operations. Then all changes to the Lagrange functions ℓ_i, $i = 1,2,\ldots,m$, are made by the updating procedure of Section 2.

We now turn to the case when the second derivative matrix of the objective function has the sparsity structure

$$d^2 F(\underline{x}) \,/\, dx_j \, dx_k = 0, \qquad \underline{x} \in \mathcal{R}^n, \quad (j,k) \in \mathcal{S}, \tag{5.8}$$

where $\mathcal{S}$ is a set of pairs of integers from $[1,n]$, such that $(j,k) \in \mathcal{S}$ implies $(k,j) \in \mathcal{S}$ and $j \neq k$. Then, for every (j,k) in $\mathcal{S}$, we set the parameter $(G_Q)_{jk}$ of the initial quadratic model to zero, so we exclude the interpolation point $\underline{x}_0 + \gamma_j \underline{e}_j + \gamma_k \underline{e}_k$ from expression (5.3). Hence the initial interpolation points are now $\underline{x}_1$ and the vectors (5.1) as before, and also all of the points (5.3) with $(j,k)\notin\mathcal{S}$. We reduce m to the new number of interpolation points, and we continue to let $\underline{x}_i$, $i=1,2,\ldots,m$, be the current set of points. Therefore, for every $(j,k)\notin\mathcal{S}$, $1\le j<k\le n$, there is a unique integer i in $[2n+2,m]$ such that the initial choice of $\underline{x}_i$ has the form $\underline{x}_0+\gamma_j\underline{e}_j+\gamma_k\underline{e}_k$ for nonzero values of γ_j and γ_k. We now let $i(j,k)$ be this value of i, so that the parameter $(G_Q)_{jk}$ of the initial quadratic model can still be obtained by equating expression (5.2) to $F(\underline{x}_{i(j,k)})$.

The initial Lagrange functions ℓ_i, $i=1,2,\ldots,m$, in the case (5.8) are as follows. We retain the quadratic polynomial (5.4) with $i=i(j,k)$ for every pair of integers (j,k) that satisfies $1\le j<k\le n$ and $(j,k)\notin\mathcal{S}$, because these polynomials not only provide the Lagrange conditions $\ell_i(\underline{x}_t)=\delta_{ti}$, $t=1,2,\ldots,m$, but also their second derivative matrices have the required sparsity structure. Then equations (5.5), (5.6) and (5.7) with the new value of m define the remaining initial Lagrange functions, which also have the sparsity structure. Further, the sparsity is preserved by the updating formulae (2.5) and (2.6). Therefore, if the known zero elements of second derivative matrices are not stored explicitly, then the work of updating all the Lagrange functions is $\mathcal{O}(m^2)$ for the new value of m, which may compare very favourably with the $\mathcal{O}(n^4)$ complexity that occurred before. We note also that the highly useful error bound (1.6) remains valid, because equation (1.4) still holds, and because the assumption $\nabla^2 F(\underline{y})=0$ in the first paragraph of the proof of Theorem 2 is in agreement with the sparsity conditions.

Other reductions in m may be possible in linearly constrained problems. We have in mind that there are fewer degrees of freedom in the variables if $\underline{x}$ is confined to the boundaries of some linear constraints. This situation arises not only when equality constraints are present but also in active set methods for inequality constraints. Therefore our work may assist the development of several new optimization algorithms that do not require the calculation of first derivatives of the objective function.

References

A.R. Conn, K. Scheinberg and Ph.L. Toint (1997a), "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 Ph.L. Toint (1997b), "Recent progress in unconstrained nonlinear optimization without derivatives", *Math. Programming*, Vol. 79, pp. 397–414.

A.R. Conn and Ph.L. Toint (1996), "An algorithm using quadratic interpolation for unconstrained derivative free optimization", in *Nonlinear Optimization* *and Applications*, eds. G. Di Pillo and F. Gianessi, Plenum Publishing (New York), pp. 27–47.

P.J. Davis (1975), *Interpolation and Approximation*, Dover Publications (New York).

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. S. Gomez and J-P. Hennart, Kluwer Academic (Dordrecht), pp. 51–67.

M.J.D. Powell (1998), "Direct search algorithms for optimization calculations", *Acta Numerica*, Vol. 7, pp. 287–336.

S. Waldron (1998), "Multipoint Taylor formulae", *Numer. Mathematik*, Vol. 80, pp. 461–494.

D. Winfield (1973), "Function minimization by interpolation in a data table", *J. Inst. Maths Applics*, Vol. 12, pp. 339–347.