DAMTP 2002/NA08

Least Frobenius norm updating of quadratic models that satisfy interpolation conditions[1]

M.J.D. Powell

Abstract: Quadratic models of objective functions are highly useful in many optimization algorithms. They are updated regularly to include new information about the objective function, such as the difference between two gradient vectors. We consider the case, however, when each model interpolates some function values, so an update is required when a new function value replaces an old one. We let the number of interpolation conditions, m say, be such that there is freedom in each new quadratic model that is taken up by minimizing the Frobenius norm of the second derivative matrix of the change to the model. This variational problem is expressed as the solution of an $(m+n+1)\times(m+n+1)$ system of linear equations, where n is the number of variables of the objective function. Further, the inverse of the matrix of the system provides the coefficients of quadratic Lagrange functions of the current interpolation problem. A method is presented for updating all these coefficients in $\mathcal{O}(\{m+n\}^2)$ operations, which allows the model to be updated too. An extension to the method is also described that suppresses the constant terms of the Lagrange functions. These techniques have a useful stability property that is investigated in some numerical experiments.

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

October, 2002 (Revised May, 2003).

[1] This paper is dedicated to Roger Fletcher, in gratitude for our collaboration, and in celebration of his 65th birthday.

1. Introduction

Let the least value of an objective function $F(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, be required, where $F(\underline{x})$ can be calculated for any vector of variables $\underline{x}\in\mathcal{R}^n$, but derivatives of F are not available. Several iterative algorithms have been developed for finding a solution to this unconstrained minimization problem, and many of them make changes to the variables that are derived from quadratic models of F. We address such algorithms, letting the current model be the quadratic polynomial

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

where $\underline{x}_0$ is a fixed vector that is often zero. On the other hand, the scalar $c\in\mathcal{R}$, the components of the vector $\underline{g}\in\mathcal{R}^n$, and the elements of the $n\times n$ matrix G, which is symmetric, are parameters of the model, whose values should be chosen so that useful accuracy is achieved in the approximation $Q(\underline{x})\approx F(\underline{x})$, if $\underline{x}$ is any candidate for the next trial vector of variables.

We see that the number of independent parameters of Q is $\frac{1}{2}(n+1)(n+2)=m^*$, say, because $\underline{x}_0$ is fixed and G is symmetric. We assume that some or all of the freedom in their values is taken up by the interpolation conditions

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

the points $\underline{x}_i$, $i=1,2,\ldots,m$, being chosen by the algorithm, and usually all the right hand sides have been calculated before starting the current iteration. We require the constraints (1.2) on the parameters of Q to be linearly independent. In other words, if $\mathcal{Q}$ is the linear space of polynomials of degree at most two from $\mathcal{R}^n$ to $\mathcal{R}$ that are zero at $\underline{x}_i$, $i=1,2,\ldots,m$, then the dimension of $\mathcal{Q}$ is m^*-m. It follows that m is at most m^*. Therefore the right hand sides of expression (1.2) are a subset of the calculated function values, if more than m^* values of the objective function were generated before the current iteration. Instead, however, all the available values of F can be taken into account by constructing quadratic models by fitting techniques, but we do not consider this subject.

We define $\underline{x}_b$ to be the best vector of variables so far, where b is an integer from $[1,m]$ that has the property

$$F(\underline{x}_b) \;=\; \min\{F(\underline{x}_i) : i=1,2,\ldots,m\}. \tag{1.3}$$

Therefore $F(\underline{x}_b)$ has been calculated, and the following method ensures that it is the least of the known function values. If the current iteration generates the new trial vector $\underline{x}^+$, if $F(\underline{x}^+)$ is calculated, and if the strict reduction $F(\underline{x}^+)<F(\underline{x}_b)$ occurs, then $\underline{x}^+$ becomes the best vector of variables, and $\underline{x}^+$ is always chosen as one of the interpolation points of the next quadratic model, Q^+ say, Otherwise, in the case $F(\underline{x}^+)\ge F(\underline{x}_b)$, the point $\underline{x}_b$ is retained as the best vector of variables and as one of the interpolation points, and it is usual, but not mandatory, to include the equation $Q^+(\underline{x}^+)=F(\underline{x}^+)$ among the constraints on Q^+.

The position of $\underline{x}_b$ is central to the choice of $\underline{x}^+$ in trust region methods. Indeed, $\underline{x}^+$ is calculated to be a sufficiently accurate estimate of the vector $\underline{x} \in \mathcal{R}^n$ that solves the subproblem

$$\text{Minimize} \quad Q(\underline{x}) \quad \text{subject to} \quad \|\underline{x} - \underline{x}_b\| \leq \rho, \tag{1.4}$$

where the norm is usually Euclidean, and where ρ is a positive parameter (namely the trust region radius), whose value is adjusted automatically. Thus $\underline{x}^+$ is bounded even if the second derivative matrix G has some negative eigenvalues. Many of the details and properties of trust region methods are studied in the books of Fletcher (1987) and of Conn, Gould and Toint (2000). Further, Conn, Scheinberg and Toint (1997) consider trust region algorithms when derivatives of the objective function F are not available. On some iterations $\underline{x}^+$ may be generated in a different way that is intended to improve the accuracy of the quadratic model.

An algorithm of this kind, namely UOBYQA, has been developed by the author (Powell, 2002), and here the interpolation conditions (1.2) define the quadratic model $Q(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, because the value of m is $m = m^* = \frac{1}{2}(n+1)(n+2)$ throughout the calculation. Therefore expression (1.2) provides an $m \times m$ system of linear equations that determines the parameters of Q. Further, on a typical iteration that adds the new interpolation condition $Q^+(\underline{x}^+) = F(\underline{x}^+)$, the interpolation points of the new quadratic model are $\underline{x}^+$ and $m-1$ of the old points $\underline{x}_i$, $i = 1, 2, \ldots, m$. Thus all the differences between the matrices of the new and the old $m \times m$ systems are confined to the t-th rows of the matrices, where $\underline{x}_t$ is the old interpolation point that is dismissed. It follows that, by applying updating techniques, the parameters of Q^+ can be calculated in $\mathcal{O}(m^2)$ computer operations, without retaining the right hand sides $F(\underline{x}_i)$, $i = 1, 2, \ldots, m$. UOBYQA also updates the coefficients of the quadratic Lagrange functions of the interpolation equations, which is equivalent to revising the inverse of the matrix of the system of equations. This approach provides several advantages (Powell 2001). In particular, in addition to the amount of work of each iteration being only $\mathcal{O}(m^2)$, the updating can be implemented in a stable way, and the availability of Lagrange functions assists the choice of the point $\underline{x}_t$ that is mentioned above.

UOBYQA is useful for calculating local solutions to unconstrained minimization problems, because the total number of evaluations of F seems to compare favourably with that of other algorithms, and high accuracy can be achieved when F is smooth. On the other hand, if the number of variables n is increased, then the amount of routine work of UOBYQA becomes prohibitive at about $n = 50$. Indeed, the value $m = m^* = \frac{1}{2}(n+1)(n+2)$ and the updating of the previous paragraph imply that the complexity of each iteration is of magnitude n^4. Further, the total number of iterations is typically $\mathcal{O}(n^2)$. Thus, for the Table 4 test problem of Powell (2003) for example, the total computation time of UOBYQA on a Sun Ultra 10 workstation increases from 20 to 1087 seconds when n is raised from 20 to 40. The routine work of many other procedures for unconstrained minimization without derivatives, however, is only $\mathcal{O}(n)$ or $\mathcal{O}(n^2)$ for each calculation of F (see Fletcher, 1987, and Powell, 1998, for instance), but the total number of function evaluations of direct search methods is often quite high, and those algorithms that approximate derivatives by differences are sensitive to lack of smoothness in the objective function.

Therefore we address the idea of constructing a quadratic model from m interpolation conditions when m is much less than m^* for large n. Let the quadratic polynomial (1.1) be the model at the beginning of the current iteration, and let the constraints on the new model

$$Q^+(\underline{x}) \;=\; c^+ + \underline{g}^{+T}(\underline{x}-\underline{x}_0) + \tfrac{1}{2}(\underline{x}-\underline{x}_0)^T G^+(\underline{x}-\underline{x}_0), \qquad \underline{x}\in\mathcal{R}^n, \tag{1.5}$$

be the equations

$$Q^+(\underline{x}_i^+) = F(\underline{x}_i^+), \qquad i=1,2,\ldots,m. \tag{1.6}$$

We take the view that Q is a useful approximation to F. Therefore, after satisfying the conditions (1.6), we employ the freedom that remains in Q^+ to minimize some measure of the difference Q^+-Q. Further, we require the change from Q to Q^+ to be independent of the choice of the fixed vector $\underline{x}_0$. Hence, because second derivative matrices of quadratic functions are independent of shifts of origin, it may be suitable to let G^+ be the $n\times n$ symmetric matrix that minimizes the square of the Frobenius norm

$$\|G^+ - G\|_F^2 \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n}(G_{ij}^+ - G_{ij})^2, \tag{1.7}$$

subject to the existence of $c^+\in\mathcal{R}$ and $\underline{g}^+\in\mathcal{R}^n$ such that the function (1.5) obeys the equations (1.6). This method defines G^+ uniquely, whenever the constraints (1.6) are consistent, because the Frobenius norm is strictly convex. Further, we assume that the corresponding values of c^+ and $\underline{g}^+$ are also unique, which imposes another condition on the positions of the interpolation points. Specifically, they must have the property that, if $p(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is any linear polynomial that satisfies $p(\underline{x}_i^+)=0$, $i=1,2,\ldots,m$, then p is identically zero. Thus m is at least $n+1$, but we require $m\geq n+2$, in order that the difference G^+-G can be nonzero.

The minimization of the Frobenius norm of the change to the second derivative matrix of the quadratic model also occurs in a well-known algorithm for unconstrained minimization when first derivatives are available, namely the symmetric Broyden method, which is described on page 73 of Fletcher (1987). There each iteration adjusts the vector of variables by a step in the space of the variables, $\underline{\delta}$ say, and the corresponding change in the gradient of the objective function, $\underline{\gamma}$ say, is calculated. The equation $\nabla^2 F\underline{\delta}=\underline{\gamma}$ would hold if F were a quadratic function. Therefore the new quadratic model (1.5) of the current iteration is given the property $G^+\underline{\delta}=\underline{\gamma}$, which corresponds to the interpolation equations (1.6), and the remaining freedom in G^+ is taken up in the way that is under consideration, namely the minimization of expression (1.7) subject to the symmetry condition ${G^+}^T = G^+$. Moreover, for the new algorithm one can form linear combinations of the constraints (1.6) that eliminate c^+ and $\underline{g}^+$, which provides $m-n-1$ independent linear constraints on the elements of G^+ that are without c^+ and $\underline{g}^+$. Thus the new updating technique is analogous to the symmetric Broyden formula.

Some preliminary experiments on applying this technique with $m=2n+1$ are reported by Powell (2003), the calculations being performed by a modified version of the UOBYQA software. The positions of the interpolation points are chosen so that the equations (1.2) would define Q if $\nabla^2 Q$ were forced to be diagonal, which is a crude way of ensuring that the equations are consistent when there are no restrictions on the symmetric matrix $\nabla^2 Q$. Further, the second derivative matrix of the first quadratic model is diagonal, but this property is not retained, because all subsequent models are constructed by the least Frobenius norm updating method that we are studying. The experiments include the solution of the Table 4 test problems of Powell (2003) to high accuracy, the ratio of the initial to the final calculated value of $F-F^*$ being about 10^{14}, where F^* is the least value of the objective function. The total numbers of evaluations of F that occurred are 2179, 4623 and 9688 in the cases $n=40$, $n=80$ and $n=160$, respectively.

These numerical results are very encouraging. In particular, when $n=160$, a quadratic model has 13041 independent parameters, so the number of function evaluations of the modified form of UOBYQA is much less than that of the usual form. Therefore high accuracy in the solution of an optimization problem may not require high accuracy in any of the quadratic models. Instead, the model should provide useful estimates of the changes to the objective function that occur for the changes to the variables that are actually made. If an estimate is poor, the discrepancy causes a substantial improvement in the model automatically, but we expect these improvements to become smaller as the iterations proceed. Indeed, it is shown in the next section that, if F is quadratic, then the least Frobenius norm updating method has the property

$$\begin{aligned} \|\nabla^2 Q^+ - \nabla^2 F\|_F^2 &= \|\nabla^2 Q - \nabla^2 F\|_F^2 - \|\nabla^2 Q^+ - \nabla^2 Q\|_F^2 \\ &\leq \|\nabla^2 Q - \nabla^2 F\|_F^2, \end{aligned} \qquad (1.8)$$

so the difference $\nabla^2 Q^+ - \nabla^2 Q$ tends to zero eventually. Therefore the construction of suitable quadratic models by the new updating technique may require fewer than $\mathcal{O}(n^2)$ function evaluations for large n, as indicated by the figures of the provisional algorithm in the last sentence of the previous paragraph. This conjecture is analogous to the important findings of Broyden, Dennis and Moré (1973) on the accuracy of second derivative estimates in gradient algorithms for unconstrained optimization.

There are now two good reasons for investigating the given updating technique. The original aim is to reduce the value of m in the systems (1.2) and (1.6) from

$m^* = \frac{1}{2}(n+1)(n+2)$ to about $2n+1$, for example, as the routine work of an iteration is at least of magnitude m^2. Secondly, the remarks of the last two paragraphs suggest that, for large n, the choice $m = m^*$ is likely to be inefficient in terms of the total number of values of the objective function that occur. Therefore the author has begun to develop software for unconstrained optimization that employs the least Frobenius norm updating procedure. The outstanding questions include the value of m, the point to remove from a set of interpolation points in order to make room for a new one, and finding a suitable method for the approximate solution of the trust region subproblem (1.4), because that task may become the most expensive part of each iteration. Here we are assuming that the updating can be implemented without serious loss of accuracy in only $\mathcal{O}(m^2)$ operations, even in the case $m = \mathcal{O}(n)$. Such implementations are studied in the remainder of this paper, in the case when every update of the set of interpolation points is the replacement of just one point by a new one, so m does not change.

In Section 2, the calculation of the new quadratic model Q^+ is expressed as the solution of an $(m+n+1) \times (m+n+1)$ system of linear equations, and the property (1.8) is established when F is quadratic. We let W^+ be the matrix of this system, and we let W be the corresponding matrix if $\underline{x}_i^+$ is replaced by $\underline{x}_i$ for $i = 1, 2, \ldots, m$. In Section 3, the inverse matrix $H = W^{-1}$ is related to the Lagrange functions of the equations (1.2), where the Frobenius norm of the second derivative matrix of each Lagrange function is as small as possible, subject to symmetry and the Lagrange conditions. Further, the usefulness of the Lagrange functions is considered, and we decide to work explicitly with the elements of H. Therefore Section 4 addresses the updating of H when just one of the points $\underline{x}_i$, $i = 1, 2, \ldots, m$, is altered. We develop a procedure that requires only $\mathcal{O}(m^2)$ operations and that has a useful stability property. The choice of $\underline{x}_0$ in expression (1.1) is also important to accuracy, but good choices are close to the optimal vector of variables, which is unknown, so it is advantageous to change $\underline{x}_0$ occasionally. That task is the subject of Section 5. Furthermore, in Section 6 the suppression of the row and column of H that holds the constant terms of the Lagrange functions is proposed, because the Lagrange conditions provide good substitutes for these terms, and the elimination of the constant terms brings some advantages. Finally, Section 7 presents and discusses numerical experiments on the stability of the given updating procedure when the number of iterations is large. They show in most cases that good accuracy is maintained throughout the calculations.

2. The solution of a variational problem

The $(m+n+1) \times (m+n+1)$ matrix W^+, mentioned in the previous paragraph, depends only on the vectors $\underline{x}_0$ and $\underline{x}_i^+$, $i = 1, 2, \ldots, m$. Therefore the same matrix would occur if the old quadratic model Q were identically zero. We begin by studying this case, and for the moment we simplify the notation by dropping the

"+" superscripts, which gives the following variational problem. It is shown later that the results of this study yield an implementation of the least Frobenius norm updating method.

We seek the quadratic polynomial (1.1) whose second derivative matrix $G = \nabla^2 Q$ has the least Frobenius norm subject to symmetry and the constraints (1.2). The vector $\underline{x}_0$, the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$, and the right hand sides $F(\underline{x}_i)$, $i=1,2,\ldots,m$, are data. It is stated in Section 1 that the positions of these points are required to have the properties:

(A1) Let $\mathcal{Q}$ be the space of quadratic polynomials from $\mathcal{R}^n$ to $\mathcal{R}$ that are zero at $\underline{x}_i$, $i=1,2,\ldots,m$. Then the dimension of $\mathcal{Q}$ is $m^* - m$, where $m^* = \frac{1}{2}(n+1)(n+2)$.

(A2) If $p(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, is any linear polynomial that is zero at $\underline{x}_i$, $i=1,2,\ldots,m$, then p is identically zero.

These properties can be achieved in many ways, and a useful technique for maintaining them when an interpolation point is moved is given in Section 3.

Condition (A1) implies that the constraints (1.2) are consistent, so we can choose a quadratic polynomial Q_0 that satisfies them. Hence the required Q has the form

$$Q(\underline{x}) \;=\; Q_0(\underline{x}) - q(\underline{x}), \qquad \underline{x} \in \mathcal{R}^n, \tag{2.1}$$

where q is the element of $\mathcal{Q}$ that gives the least value of the Frobenius norm $\|\nabla^2 Q_0 - \nabla^2 q\|_F$. This condition provides a unique matrix $\nabla^2 q$. Moreover, if two different functions $q \in \mathcal{Q}$ have the same second derivative matrix, then the difference between them is a nonzero linear polynomial, which is not allowed by condition (A2). Therefore the given variational problem has a unique solution of the form (1.1).

Next we identify a useful system of linear equations that provides the parameters $c \in \mathcal{R}$, $\underline{g} \in \mathcal{R}^n$ and $G \in \mathcal{R}^{n \times n}$ of this solution. We deduce from the equations (1.1) and (1.2) that the parameters minimize the function

$$\tfrac{1}{4}\,\|G\|_F^2 \;=\; \tfrac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ij}^2, \tag{2.2}$$

subject to the linear constraints

$$c + \underline{g}^T(\underline{x}_i - \underline{x}_0) + \tfrac{1}{2}(\underline{x}_i - \underline{x}_0)^T G\,(\underline{x}_i - \underline{x}_0) \;=\; F(\underline{x}_i), \qquad i=1,2,\ldots,m, \tag{2.3}$$

and $G^T = G$, which is a convex quadratic programming problem. We drop the condition that G be symmetric, however, because without it the symmetry of G occurs automatically. Therefore there exist Lagrange multipliers λ_k, $k=1,2,\ldots,m$, such that the first derivatives of the expression

$$L(c,\underline{g},G) \;=\; \tfrac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ij}^2 - \sum_{k=1}^{m} \lambda_k \left\{ c + \underline{g}^T(\underline{x}_k - \underline{x}_0) + \tfrac{1}{2}(\underline{x}_k - \underline{x}_0)^T G\,(\underline{x}_k - \underline{x}_0) \right\}, \tag{2.4}$$

with respect to the parameters of Q, are all zero at the solution of the quadratic programming problem. In other words, the Lagrange multipliers and the required values of the parameters satisfy the equations

$$\left.\begin{array}{l} \sum_{k=1}^m \lambda_k = 0, \qquad \sum_{k=1}^m \lambda_k \, (\underline{x}_k - \underline{x}_0) = 0 \\ \text{and} \quad G \;=\; \sum_{k=1}^m \lambda_k \, (\underline{x}_k - \underline{x}_0)(\underline{x}_k - \underline{x}_0)^T \end{array}\right\}. \tag{2.5}$$

The second line of this expression shows the symmetry of G, and is derived by differentiating the function (2.4) with respect to the elements of G, while the two equations in the first line are obtained by differentiation with respect to c and the components of $\underline{g}$. Now first order conditions are necessary and sufficient for optimality in convex optimization calculations (see Theorem 9.4.2 of Fletcher, 1987). Further, we have found already that the required parameters are unique, and the Lagrange multipliers at the solution of the quadratic programming problem are also unique, because the constraints (2.3) are linearly independent. It follows that the values of all these parameters and multipliers are defined by the equations (2.3) and (2.5).

We use the second line of expression (2.5) to eliminate G from these equations. Thus the constraints (2.3) take the form

$$c + \underline{g}^T (\underline{x}_i - \underline{x}_0) + \tfrac{1}{2} \sum_{k=1}^m \lambda_k \, \{(\underline{x}_i - \underline{x}_0)^T (\underline{x}_k - \underline{x}_0)\}^2 \;=\; F(\underline{x}_i), \quad i=1,2,\ldots,m. \tag{2.6}$$

We let A be the $m \times m$ matrix that has the elements

$$A_{ik} \;=\; \tfrac{1}{2} \, \{(\underline{x}_i - \underline{x}_0)^T (\underline{x}_k - \underline{x}_0)\}^2, \qquad 1 \le i,k \le m, \tag{2.7}$$

we let $\underline{e}$ and $\underline{F}$ be the vectors in $\mathcal{R}^m$ whose components are $e_i = 1$ and $F_i = F(\underline{x}_i)$, $i=1,2,\ldots,m$, and we let X be the $n \times m$ matrix whose columns are the differences $\underline{x}_k - \underline{x}_0$, $k=1,2,\ldots,m$. Thus the conditions (2.6) and the first line of expression (2.5) give the $(m+n+1) \times (m+n+1)$ system of equations

$$\left(\begin{array}{c|c} A & \underline{e} \;\vdots\; X^T \\ \hline \underline{e}^T & \\ \cdots & 0 \\ X & \end{array}\right) \left(\begin{array}{c} \underline{\lambda} \\ \hline c \\ \cdots \\ \underline{g} \end{array}\right) \;=\; W \left(\begin{array}{c} \underline{\lambda} \\ \hline c \\ \cdots \\ \underline{g} \end{array}\right) \;=\; \left(\begin{array}{c} \underline{F} \\ \hline 0 \\ \cdots \\ \underline{0} \end{array}\right), \tag{2.8}$$

where W is introduced near the end of Section 1, and is nonsingular because of the last remark of the previous paragraph.

We see that W is symmetric. We note also that its leading $m \times m$ submatrix, namely A, has no negative eigenvalues, which is proved by establishing $\underline{v}^T A \, \underline{v} \ge 0$, where $\underline{v}$ is any vector in $\mathcal{R}^m$. Specifically, because the definitions of A and X provide the formula

$$A_{ik} \;=\; \tfrac{1}{2} \, \{(\underline{x}_i - \underline{x}_0)^T (\underline{x}_k - \underline{x}_0)\}^2 \;=\; \tfrac{1}{2} \Big\{ \sum_{s=1}^n X_{si} X_{sk} \Big\}^2, \quad 1 \le i,k \le m, \tag{2.9}$$

we find the required inequality

$$\begin{aligned}\underline{v}^T A\, \underline{v} \;&=\; \tfrac{1}{2} \sum_{i=1}^{m} \sum_{k=1}^{m} \sum_{s=1}^{n} \sum_{t=1}^{n} v_i\, v_k\, X_{si}\, X_{sk}\, X_{ti}\, X_{tk} \\ &=\; \tfrac{1}{2} \sum_{s=1}^{n} \sum_{t=1}^{n} \Big\{ \sum_{i=1}^{m} v_i\, X_{si}\, X_{ti} \Big\}^2 \;\geq\; 0. \end{aligned} \tag{2.10}$$

Moreover, for any fixed vector $\underline{x}_0$, condition (A2) at the beginning of this section is equivalent to the linear independence of the last $n+1$ rows or columns of W.

We now turn our attention to the updating calculation of Section 1. The new quadratic model (1.5) is constructed by minimizing the Frobenius norm of the second derivative matrix of the difference

$$(Q^+ - Q)(\underline{x}) \;=\; c^\# + \underline{g}^{\#T} (\underline{x} - \underline{x}_0) + \tfrac{1}{2} (\underline{x} - \underline{x}_0)^T G^\# (\underline{x} - \underline{x}_0), \qquad \underline{x} \in \mathcal{R}^n, \tag{2.11}$$

subject to the constraints

$$(Q^+ - Q)(\underline{x}_i^+) \;=\; F(\underline{x}_i^+) - Q(\underline{x}_i^+), \qquad i = 1, 2, \ldots, m, \tag{2.12}$$

the variables of this calculation being $c^\# \in \mathcal{R}$, $\underline{g}^\# \in \mathcal{R}^n$ and $G^\# \in \mathcal{R}^{n \times n}$. This variational problem is the one we have studied already, if we replace expressions (1.1) and (1.2) by expressions (2.11) and (2.12), respectively, and if we alter the interpolation points in conditions (A1) and (A2) from $\underline{x}_i$ to $\underline{x}_i^+$, $i = 1, 2, \ldots, m$. Therefore the analogue of the system (2.8), whose matrix is called W^+ near the end of Section 1, defines the quadratic polynomial $Q^+ - Q$, which is added to Q, in order to generate Q^+. A convenient form of this procedure is presented later, which takes advantage of the assumption that every update of the set of interpolation points is the replacement of just one point by a new one. If $\underline{x}_i^+$ is in the set $\{\underline{x}_j : j = 1, 2, \ldots, m\}$, then the conditions (1.2) on Q imply that the right hand side of expression (2.12) is zero. It follows that at most one of the constraints (2.12) on the difference $Q^+ - Q$ has a nonzero right hand side. Thus the Lagrange functions of the next section become highly useful.

The proof of the assertion (1.8) when F is quadratic is elementary. Specifically, we let Q^+ be given by the method of the previous paragraph, where the interpolation points can have any positions that are allowed by conditions (A1) and (A2). Further, we let θ be any real number, and we consider the function $\{(Q^+(\underline{x}) - Q(\underline{x})\} + \theta \{F(\underline{x}) - Q^+(\underline{x})\}$, $\underline{x} \in \mathcal{R}^n$. It is a quadratic polynomial, and its values at $\underline{x}_i^+$, $i = 1, 2, \ldots, m$, are independent of θ, because of the conditions (1.6) on Q^+. It follows from the given construction of $Q^+ - Q$ that the least value of the Frobenius norm

$$\| (\nabla^2 Q^+ - \nabla^2 Q) + \theta\, (\nabla^2 F - \nabla^2 Q^+) \|_F, \qquad \theta \in \mathcal{R}, \tag{2.13}$$

occurs when θ is zero, which implies the equation

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \Big\{ (\nabla^2 Q^+)_{ij} - (\nabla^2 Q)_{ij} \Big\} \Big\{ (\nabla^2 F)_{ij} - (\nabla^2 Q^+)_{ij} \Big\} \;=\; 0. \tag{2.14}$$

We see that the left hand side of this identity is half the difference between the right and left hand sides of the first line of expression (1.8). Therefore the properties (1.8) are achieved. They show that, if F is quadratic, then the sequence of iterations causes $\|\nabla^2 Q - \nabla^2 F\|_F$ and $\|\nabla^2 Q^+ - \nabla^2 Q\|_F$ to decrease monotonically and to tend to zero, respectively.

3. The Lagrange functions of the interpolation equations

From now on, the meaning of the term Lagrange function is taken from polynomial interpolation instead of from the theory of constrained optimization. Specifically, the Lagrange functions of the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$, are quadratic polynomials $\ell_j(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, $j=1,2,\ldots,m$, that satisfy the conditions

$$\ell_j(\underline{x}_i) \;=\; \delta_{ij}, \qquad 1\le i,j\le m, \tag{3.1}$$

where δ_{ij} is the Kronecker delta. Further, in order that they are applicable to the variational problem of Section 2, we retain the conditions (A1) and (A2) on the positions of the interpolation points, and, for each j, we take up the freedom in ℓ_j by minimizing the Frobenius norm $\|\nabla^2\ell_j\|_F$, subject to the constraints (3.1). Therefore the parameters of ℓ_j are defined by the linear system (2.8), if we replace the right hand side of this system by the j-th coordinate vector in $\mathcal{R}^{m+n+1}$. Thus, if we let Q be the quadratic polynomial

$$Q(\underline{x}) \;=\; \sum_{j=1}^{m} F(\underline{x}_j)\,\ell_j(\underline{x}), \qquad \underline{x}\in\mathcal{R}^n, \tag{3.2}$$

then its parameters satisfy the given equations (2.8). It follows from the nonsingularity of this system of equations that expression (3.2) is the Lagrange form of the solution of the variational problem of Section 2.

Let H be the inverse of the matrix W of the system (2.8), as stated in the last paragraph of Section 1. The given definition of ℓ_j, where j is any integer from $[1,m]$, implies that the j-th column of H provides the parameters of ℓ_j. In particular, because of the second line of expression (2.5), ℓ_j has the second derivative matrix

$$G_j \;=\; \nabla^2\ell_j \;=\; \sum_{k=1}^{m} H_{kj}\,(\underline{x}_k-\underline{x}_0)(\underline{x}_k-\underline{x}_0)^T, \qquad j=1,2,\ldots,m. \tag{3.3}$$

Further, letting c_j and $\underline{g}_j$ be $H_{m+1\,j}$ and the vector in $\mathcal{R}^n$ with components H_{ij}, $i=m+2,m+3,\ldots,m+n+1$, respectively, we find that ℓ_j is the polynomial

$$\ell_j(\underline{x}) \;=\; c_j + \underline{g}_j^T(\underline{x}-\underline{x}_0) + \tfrac{1}{2}(\underline{x}-\underline{x}_0)^T G_j\,(\underline{x}-\underline{x}_0), \qquad \underline{x}\in\mathcal{R}^n. \tag{3.4}$$

Because the Lagrange functions occur explicitly in some of the techniques of the optimization software, we require the elements of H to be available, but there is no need to store the matrix W.

Let $\underline{x}^+$ be the new vector of variables, as introduced in the paragraph that includes expression (1.4). In the usual case when $\underline{x}^+$ replaces one of the points $\underline{x}_i$, $i=1,2,\ldots,m$, we let $\underline{x}_t$ be dismissed, so the new interpolation points are the vectors

$$\underline{x}_t^+=\underline{x}^+ \quad \text{and} \quad \underline{x}_i^+=\underline{x}_i, \quad i\in\{1,2,\ldots,m\}\backslash\{t\}. \tag{3.5}$$

One advantage of the Lagrange functions is that they provide a convenient way of maintaining the conditions (A1) and (A2). Indeed, it is shown below that these conditions are inherited by the new interpolation points if t is chosen so that $\ell_t(\underline{x}^+)$ is nonzero. All of the numbers $\ell_j(\underline{x}^+)$, $j=1,2,\ldots,m$, can be generated in only $\mathcal{O}(m^2)$ operations when H is available, by first calculating the scalar products

$$\sigma_k \;=\; (\underline{x}_k-\underline{x}_0)^T(\underline{x}^+-\underline{x}_0), \qquad k=1,2,\ldots,m, \tag{3.6}$$

and then applying the formula

$$\ell_j(\underline{x}^+) \;=\; c_j+\underline{g}_j^T(\underline{x}^+-\underline{x}_0)+\tfrac{1}{2}\sum_{k=1}^{m} H_{kj}\,\sigma_k^2, \qquad j=1,2,\ldots,m, \tag{3.7}$$

which is derived from equations (3.3) and (3.4). At least one of the numbers (3.7) is nonzero, because interpolation to a constant function yields the identity

$$\sum_{j=1}^{m} \ell_j(\underline{x}) \;=\; 1, \qquad \underline{x}\in\mathcal{R}^n. \tag{3.8}$$

Let $\ell_t(\underline{x}^+)$ be nonzero, let condition (A1) at the beginning of Section 2 be satisfied, and let $\mathcal{Q}^+$ be the space of quadratic polynomials from $\mathcal{R}^n$ to $\mathcal{R}$ that are zero at $\underline{x}_i^+$, $i=1,2,\ldots,m$. We have to prove that the dimension of Q^+ is m^*-m. We employ the linear space, $\mathcal{Q}^-$ say, of quadratic polynomials that are zero at $\underline{x}_i^+=\underline{x}_i$, $i\in\{1,2,\ldots,m\}\backslash\{t\}$. It follows from condition (A1) that the dimension of $\mathcal{Q}^-$ is m^*-m+1. Further, the dimension of $\mathcal{Q}^+$ is m^*-m if and only if an element of $\mathcal{Q}^-$ is nonzero at $\underline{x}_t^+=\underline{x}^+$. The Lagrange equations (3.1) show that ℓ_t is in $\mathcal{Q}^-$. Therefore the property $\ell_t(\underline{x}^+)\neq 0$ gives the required result.

We now consider condition (A2). It is achieved by the new interpolation points if the values

$$p(\underline{x}_i) \;=\; 0, \qquad i\in\{1,2,\ldots,m\}\backslash\{t\}, \tag{3.9}$$

where p is a linear polynomial, imply $p\equiv 0$. Otherwise, we let p be a nonzero polynomial of this kind, and we deduce from condition (A2) that $p(\underline{x}_t)$ is nonzero. Therefore, because all second derivatives of p are zero, the function $p(\underline{x})/p(\underline{x}_t)$, $\underline{x}\in\mathcal{R}^n$, is the Lagrange function ℓ_t. Thus, if p is a nonzero linear polynomial that takes the values (3.9), then it is a multiple of ℓ_t. Such polynomials cannot vanish at $\underline{x}_t^+$ because of the property $\ell_t(\underline{x}^+)\neq 0$. It follows that condition (A2) is also inherited by the new interpolation points.

These remarks suggest that, in the presence of computer rounding errors, the preservation of conditions (A1) and (A2) by the sequence of iterations may be more stable if $|\ell_t(\underline{x}^+)|$ is relatively large. The UOBYQA software of Powell (2002) follows this strategy when it tries to improve the accuracy of the quadratic model, which is the alternative to solving the trust region subproblem, as mentioned at the end of the paragraph that includes expression (1.4). Then the interpolation point that is going to be replaced by $\underline{x}^+$, namely $\underline{x}_t$, is selected before the position of $\underline{x}^+$ is chosen. Indeed, $\underline{x}_t$ is often the element of the set $\{\underline{x}_i : i=1,2,\ldots,m\}$ that is furthest from the best point $\underline{x}_b$, because Q is intended to be an adequate approximation to F within the trust region of subproblem (1.4). Having picked the index t, the value of $|\ell_t(\underline{x}^+)|$ is made relatively large, by letting $\underline{x}^+$ be an estimate of the vector $\underline{x}\in\mathcal{R}^n$ that solves the alternative subproblem

$$\text{Maximize} \quad |\ell_t(\underline{x})| \quad \text{subject to} \quad \|\underline{x}-\underline{x}_b\| \le \rho, \tag{3.10}$$

so again the availability of the Lagrange functions is required.

Let H and H^+ be the inverses of W and W^+, where W and W^+ are the matrices of the system (2.8) for the old and new interpolation points, respectively. The construction of the new quadratic model $Q^+(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, is going to depend on H^+. Expression (3.5), the definition (2.7) of A, and the definition of X a few lines later, imply that the differences between W and W^+ occur only in the t-th rows and columns of these matrices. Therefore the ranks of the matrices W^+-W and H^+-H are at most two. It follows that H^+ can be generated from H in only $\mathcal{O}(m^2)$ computer operations. That task is addressed in Section 4, so we assume until then that we are able to find all the elements of H^+ before beginning the calculation of Q^+.

We recall from the penultimate paragraph of Section 2 that the new model Q^+ is formed by adding the difference Q^+-Q to Q, where Q^+-Q is the quadratic polynomial whose second derivative matrix has the least Frobenius norm subject to the constraints (2.12). Further, equations (1.2) and (3.5) imply that only the t-th right hand side of these constraints can be nonzero. Therefore, by considering the Lagrange form (3.2) of the solution of the variational problem of Section 2, we deduce that Q^+-Q is a multiple of the t-th Lagrange function, ℓ_t^+ say, of the new interpolation points, where the multiplying factor is defined by the constraint (2.12) in the case $i=t$. Thus Q^+ is the quadratic

$$Q^+(\underline{x}) \;=\; Q(\underline{x}) + \{F(\underline{x}^+)-Q(\underline{x}^+)\}\,\ell_t^+(\underline{x}), \qquad \underline{x}\in\mathcal{R}^n. \tag{3.11}$$

Moreover, by applying the techniques in the second paragraph of this section, the values of all the parameters of ℓ_t^+ are deduced from the elements of the t-th column of H^+. It follows that the constant term c^+ and the components of the vector $\underline{g}^+$ of the new model (1.5) are the sums

$$\left.\begin{array}{rcl} c^+ &=& c + \{F(\underline{x}^+)-Q(\underline{x}^+)\}\,H^+_{m+1\,t} \\ g_j^+ &=& g_j + \{F(\underline{x}^+)-Q(\underline{x}^+)\}\,H^+_{m+j+1\,t}, \quad j=1,2,\ldots,n \end{array}\right\}. \tag{3.12}$$

On the other hand, we find below that the calculation of all the elements of the second derivative matrix $G^+ = \nabla^2 Q^+$ is relatively expensive.

Formula (3.11) shows that G^+ is the matrix

$$\begin{aligned} G^+ &= G + \{F(\underline{x}^+) - Q(\underline{x}^+)\}\, \nabla^2 \ell_t^+(\underline{x}^+) \\ &= G + \{F(\underline{x}^+) - Q(\underline{x}^+)\} \sum_{k=1}^{m} H_{kt}^+\, (\underline{x}_k^+ - \underline{x}_0)(\underline{x}_k^+ - \underline{x}_0)^T, \end{aligned} \tag{3.13}$$

where the last line is obtained by setting $j=t$ in the version of expression (3.3) for the new interpolation points. We see that G^+ can be constructed by adding m matrices of rank one to G, but the work of that task would be $\mathcal{O}(mn^2)$, which is unwelcome in the case $m = \mathcal{O}(n)$, because we are trying to complete the updating in only $\mathcal{O}(m^2)$ operations. Therefore, instead of storing G explicitly, we employ the form

$$G = \Gamma + \sum_{k=1}^{m} \gamma_k\, (\underline{x}_k - \underline{x}_0)(\underline{x}_k - \underline{x}_0)^T, \tag{3.14}$$

which defines the matrix Γ for any choice of γ_k, $k=1,2,\ldots,m$, these multipliers being stored. We seek a similar expression for G^+. Specifically, because of the change (3.5) to the positions of the interpolation points, we let Γ^+ and G^+ be the matrices

$$\left. \begin{aligned} \Gamma^+ &= \Gamma + \gamma_t\, (\underline{x}_t - \underline{x}_0)(\underline{x}_t - \underline{x}_0)^T \\ G^+ &= \Gamma^+ + \sum_{k=1}^{m} \gamma_k^+\, (\underline{x}_k^+ - \underline{x}_0)(\underline{x}_k^+ - \underline{x}_0)^T \end{aligned} \right\}. \tag{3.15}$$

Then equations (3.13) and (3.14) provide the values

$$\gamma_k^+ = \gamma_k\, (1 - \delta_{kt}) + \{F(\underline{x}^+) - Q(\underline{x}^+)\}\, H_{kt}^+, \qquad k=1,2,\ldots,m, \tag{3.16}$$

where δ_{kt} is still the Kronecker delta. Thus, by expressing $G = \nabla^2 Q$ in the form (3.14), the construction of Q^+ from Q requires at most $\mathcal{O}(m^2)$ operations, which meets the target that has been mentioned. The quadratic model of the first iteration is calculated from the interpolation conditions (1.2) by solving the variational problem of Section 2. Therefore, because of the second line of expression (2.5), the choices $\Gamma = 0$ and $\gamma_k = \lambda_k$, $k=1,2,\ldots,m$, can be made initially for the second derivative matrix (3.14).

This form of G is less convenient than G itself. Fortunately, however, the work of multiplying a general vector $\underline{v} \in \mathcal{R}^n$ by the matrix (3.14) is only $\mathcal{O}(mn)$. Therefore, when developing Fortran software for unconstrained optimization that includes the least Frobenius norm updating technique, the author expects to generate an approximate solution of the trust region subproblem (1.4) by a version of the conjugate gradient method. For example, one of the procedures that are studied in Chapter 7 of Conn, Gould and Toint (2000) may be suitable.

4. The updating of the inverse matrix H

We introduce the calculation of H^+ from H by identifying the stability property that is achieved. We recall that the change (3.5) to the interpolation points causes the symmetric matrices $W = H^{-1}$ and $W^+ = (H^+)^{-1}$ to differ only in their t-th rows and columns. We recall also that W is not stored. Therefore our formula for H^+ is going to depend only on H and on the vector $\underline{w}_t^+ \in \mathcal{R}^{m+n+1}$, which is the t-th column of W^+. These data define H^+, because in theory the updating calculation can begin by inverting H to give W. Then the availability of $\underline{w}_t^+$ allows the symmetric matrix W^+ to be formed from W. Finally, H^+ is set to the inverse of W^+. This procedure provides excellent protection against the accumulation of computer rounding errors.

We are concerned about the possibility of large errors in H, due to the addition and magnification of the effects of rounding errors by a long sequence of previous iterations. Therefore, because our implementation of the calculation of H^+ from H and $\underline{w}_t^+$ is going to require only $\mathcal{O}(m^2)$ operations, we assume that the contributions to H^+ from the errors of the current iteration are negligible. On the other hand, most of the errors in H are inherited to some extent by H^+. Fortunately, we find below that this process is without growth, for a particular measure of the error in H, namely the size of the elements of $\Delta = W - H^{-1}$, where W is still the true matrix of the system (2.8). We let Δ be nonzero due to the work of previous iterations, but, as mentioned already, we ignore the new errors of the current iteration. We relate $\Delta^+ = W^+ - (H^+)^{-1}$ to Δ, where W^+ is the true matrix of the system (2.8) for the new interpolation points. It follows from the construction of the previous paragraph, where the t-th column of $(H^+)^{-1}$ is $\underline{w}_t^+$, that all elements in the t-th row and column of Δ^+ are zero. Moreover, if i and j are any integers from the set $\{1, 2, \ldots, m+n+1\} \backslash \{t\}$, then the definitions of W and W^+ imply $W_{ij}^+ = W_{ij}$, while the construction of H^+ implies $(H^+)_{ij}^{-1} = H_{ij}^{-1}$. Thus the assumptions give the property

$$\Delta_{ij}^+ \;=\; \delta_{it}\, \delta_{jt}\, \Delta_{ij}, \qquad 1 \le i, j \le m+n+1, \tag{4.1}$$

δ_{it} and δ_{jt} being the Kronecker delta. In practice, therefore, any growth of the form $|\Delta_{ij}^+| > |\Delta_{ij}|$ is due to the rounding errors of the current iteration. Further, any cumulative effects of errors in the t-th row and column of Δ are eliminated by the updating procedure, where t is the index of the new interpolation point. Some numerical experiments on these stability properties are reported in Section 7.

Two formulae for H^+ will be presented. The first one can be derived in several ways from the construction of H^+ described above. Probably the author's algebra is unnecessarily long, because it introduces a factor into a denominator that is removed algebraically. Therefore the details of that derivation are suppressed. They provide the symmetric matrix

$$H^+ \;=\; H + \frac{1}{\sigma_t^+} \Big[\alpha_t^+ \,(\underline{e}_t - H\, \underline{w}_t^+)\, (\underline{e}_t - H\, \underline{w}_t^+)^T \,-\, \beta_t^+ \, H\, \underline{e}_t\, \underline{e}_t^T H$$

$$+\,\tau_t^+\left\{H\,\underline{e}_t\,(\underline{e}_t-H\,\underline{w}_t^+)^T+(\underline{e}_t-H\,\underline{w}_t^+)\,\underline{e}_t^T H\right\}\Big], \tag{4.2}$$

where $\underline{e}_t$ is the t-th coordinate vector in $\mathcal{R}^{n+m+1}$, and where its parameters have the values

$$\left.\begin{array}{lll} \alpha_t^+ \;=\; \underline{e}_t^T H\,\underline{e}_t, & & \beta_t^+ \;=\; (\underline{e}_t-H\,\underline{w}_t^+)^T\underline{w}_t^+, \\ \tau_t^+ \;=\; \underline{e}_t^T H\,\underline{w}_t^+, & \text{and} & \sigma_t^+ \;=\; \alpha_t^+\beta_t^+ + {\tau_t^+}^2. \end{array}\right\} \tag{4.3}$$

The correctness of expression (4.2) is established in the theorem below. We see that H^+ can be calculated from H and $\underline{w}_t^+$ in only $\mathcal{O}(m^2)$ operations. The other formula for H^+, given later, has the advantage that, by making suitable changes to the parameters (4.3), $\underline{w}_t^+$ is replaced by a vector that is independent of t.

Theorem: If H is nonsingular and symmetric, and if σ_t^+ is nonzero, then expressions (4.2) and (4.3) provide the matrix H^+ that is defined in the first paragraph of this section.

Proof: H^+ is defined to be the inverse of the symmetric matrix whose t-th column is $\underline{w}_t^+$ and whose other columns are the vectors

$$\underline{v}_j = H^{-1}\underline{e}_j + \left(\underline{e}_j^T\underline{w}_t^+ - \underline{e}_j^T H^{-1}\underline{e}_t\right)\underline{e}_t, \qquad j\in\{1,2,\ldots,n{+}m{+}1\}\backslash\{t\}. \tag{4.4}$$

Therefore, letting H^+ be the matrix (4.2), it is sufficient to establish $H^+\underline{w}_t^+=\underline{e}_t$ and $H^+\underline{v}_j=\underline{e}_j$, $j\neq t$. Because equation (4.3) shows that β_t^+ and τ_t^+ are the scalar products $(\underline{e}_t-H\underline{w}_t^+)^T\underline{w}_t^+$ and $\underline{e}_t^T H\underline{w}_t^+$, respectively, formula (4.2) achieves the condition

$$\begin{aligned} H^+\underline{w}_t^+ \;&=\; H\,\underline{w}_t^+ + (\sigma_t^+)^{-1}\Big[\alpha_t^+\,\beta_t^+\,(\underline{e}_t-H\,\underline{w}_t^+) - \beta_t^+\,\tau_t^+\,H\,\underline{e}_t \\ &\qquad\qquad + \tau_t^+\left\{\beta_t^+ H\,\underline{e}_t + \tau_t^+\,(\underline{e}_t-H\,\underline{w}_t^+)\right\}\Big] \\ &=\; H\,\underline{w}_t^+ + (\sigma_t^+)^{-1}\{\alpha_t^+\,\beta_t^+ + {\tau_t^+}^2\}\,(\underline{e}_t-H\,\underline{w}_t^+) \;=\; \underline{e}_t, \end{aligned} \tag{4.5}$$

the last equation being due to the definition (4.3) of σ_t^+. It follows that, if j is any integer from $[1,n{+}m{+}1]$ that is different from t, then it remains to prove $H^+\underline{v}_j=\underline{e}_j$.

Formula (4.2), $j\neq t$ and the symmetry of H^{-1} provide the identity

$$H^+\,(H^{-1}\underline{e}_j) \;=\; \underline{e}_j + \frac{(\underline{e}_t-H\,\underline{w}_t^+)^T H^{-1}\underline{e}_j}{\sigma_t^+}\,\left[\alpha_t^+\,(\underline{e}_t-H\,\underline{w}_t^+) + \tau_t^+ H\,\underline{e}_t\right]. \tag{4.6}$$

Moreover, because the scalar products $(\underline{e}_t-H\underline{w}_t^+)^T\underline{e}_t$ and $\underline{e}_t^T H\underline{e}_t$ take the values $1-\tau_t^+$ and α_t^+, formula (4.2) also gives the property

$$\begin{aligned} H^+\,\underline{e}_t \;&=\; H\,\underline{e}_t + (\sigma_t^+)^{-1}\Big[\alpha_t^+\,(1-\tau_t^+)\,(\underline{e}_t-H\,\underline{w}_t^+) - \beta_t^+\,\alpha_t^+\,H\,\underline{e}_t \\ &\qquad\qquad + \tau_t^+\left\{(1-\tau_t^+)\,H\,\underline{e}_t + \alpha_t^+\,(\underline{e}_t-H\,\underline{w}_t^+)\right\}\Big] \\ &=\; (\sigma_t^+)^{-1}\left[\alpha_t^+\,(\underline{e}_t-H\,\underline{w}_t^+) + \tau_t^+ H\,\underline{e}_t\right]. \end{aligned} \tag{4.7}$$

The numerator in expression (4.6) has the value $-(\underline{e}_j^T \underline{w}_t^+ - \underline{e}_j^T H^{-1} \underline{e}_t)$. Therefore equations (4.4), (4.6) and (4.7) imply the condition $H^+ \underline{v}_j = \underline{e}_j$, which completes the proof. $\Box$

The vector $\underline{w}_t^+$ of formula (4.2) is the t-th column of the matrix of the system (2.8) for the new interpolation points. Therefore, because of the choice $\underline{x}_t^+ = \underline{x}^+$, it has the components

$$\left.\begin{array}{rl} (\underline{w}_t^+)_i \;=\; \tfrac{1}{2}\{(\underline{x}_i^+ - \underline{x}_0)^T (\underline{x}^+ - \underline{x}_0)\}^2, & i=1,2,\ldots,m, \\ (\underline{w}_t^+)_{m+1} \;=\; 1, \quad \text{and} \quad (\underline{w}_t^+)_{m+i+1} \;=\; (\underline{x}^+ - \underline{x}_0)_i, & i=1,2,\ldots,n. \end{array}\right\} \tag{4.8}$$

Moreover, we let $\underline{w} \in \mathcal{R}^{m+n+1}$ have the components

$$\left.\begin{array}{rl} w_i \;=\; \tfrac{1}{2}\{(\underline{x}_i - \underline{x}_0)^T (\underline{x}^+ - \underline{x}_0)\}^2, & i=1,2,\ldots,m, \\ w_{m+1} \;=\; 1, \quad \text{and} \quad w_{m+i+1} \;=\; (\underline{x}^+ - \underline{x}_0)_i, & i=1,2,\ldots,n. \end{array}\right\} \tag{4.9}$$

It follows from the positions (3.5) of the new interpolation points that $\underline{w}_t^+$ is the sum

$$\underline{w}_t^+ \;=\; \underline{w} + \eta_t \, \underline{e}_t, \tag{4.10}$$

where $\underline{e}_t$ is still the t-th coordinate vector in $\mathcal{R}^{m+n+1}$, and where η_t is the difference

$$\eta_t \;=\; \underline{e}_t^T \underline{w}_t^+ - \underline{e}_t^T \underline{w} \;=\; \tfrac{1}{2} \, \|\underline{x}^+ - \underline{x}_0\|^4 - \underline{e}_t^T \underline{w}. \tag{4.11}$$

An advantage of working with $\underline{w}$ instead of with $\underline{w}_t^+$ is that, if $\underline{x}^+$ is available before t is selected, which happens when $\underline{x}^+$ is calculated from the trust region subproblem (1.4), then $\underline{w}$ is independent of t. Therefore we derive a new version of the updating formula (4.2) by making the substitution (4.10).

Specifically, we replace $\underline{e}_t - H \underline{w}_t^+$ by $\underline{e}_t - H \underline{w} - \eta_t H \underline{e}_t$ in equation (4.2). Then some elementary algebra gives the expression

$$\begin{aligned} H^+ \;=\; H + \frac{1}{\sigma_t} \Big[& \alpha_t \, (\underline{e}_t - H \underline{w}) \, (\underline{e}_t - H \underline{w})^T - \beta_t \, H \underline{e}_t \, \underline{e}_t^T H \\ & + \tau_t \Big\{ H \underline{e}_t \, (\underline{e}_t - H \underline{w})^T + (\underline{e}_t - H \underline{w}) \, \underline{e}_t^T H \Big\} \Big], \end{aligned} \tag{4.12}$$

its parameters having the values

$$\left.\begin{array}{ll} \alpha_t \;=\; \alpha_t^+, & \beta_t \;=\; \beta_t^+ - \alpha_t^+ \eta_t^2 + 2\,\tau_t^+ \eta_t, \\ \tau_t \;=\; \tau_t^+ - \alpha_t^+ \eta_t, & \text{and} \quad \sigma_t \;=\; \sigma_t^+. \end{array}\right\} \tag{4.13}$$

The following remarks remove the "+" superscripts from these right hand sides. The definitions (4.13) imply the identity $\alpha_t \beta_t + \tau_t^2 = \alpha_t^+ \beta_t^+ + {\tau_t^+}^2$, so expression (4.3) with $\alpha_t = \alpha_t^+$ provides the formulae

$$\alpha_t \;=\; \underline{e}_t^T H \underline{e}_t \qquad \text{and} \qquad \sigma_t \;=\; \alpha_t \beta_t + \tau_t^2. \tag{4.14}$$

Further, by combining equation (4.10) with the values (4.3), we deduce the forms

$$\begin{aligned}\beta_t &= (\underline{e}_t - H\underline{w} - \eta_t H\underline{e}_t)^T(\underline{w}+\eta_t\underline{e}_t) - \eta_t^2\,\underline{e}_t^T H\underline{e}_t + 2\,\eta_t\,\underline{e}_t^T H(\underline{w}+\eta_t\underline{e}_t)\\ &= (\underline{e}_t - H\underline{w})^T\underline{w} + \eta_t, \qquad \text{and} \qquad\qquad (4.15)\\ \tau_t &= \underline{e}_t^T H(\underline{w}+\eta_t\underline{e}_t) - \eta_t\,\underline{e}_t^T H\underline{e}_t = \underline{e}_t^T H\underline{w}. \qquad\qquad (4.16)\end{aligned}$$

It is straightforward to verify that equations (4.12) and (4.14)–(4.16) give the property $H^+(\underline{w}+\eta_t\underline{e}_t)=\underline{e}_t$, which is equivalent to condition (4.5).

Another advantage of working with $\underline{w}$ instead of with $\underline{w}_t^+$ in the updating procedure is that the first m components of the product $H\underline{w}$ are the values $\ell_j(\underline{x}^+)$, $j=1,2,\ldots,m$, of the current Lagrange functions at the new point $\underline{x}^+$. We justify this assertion by recalling equations (3.3) and (3.4), and the observation that the elements $H_{m+1\,j}$ and H_{ij}, $i=m+2,m+3,\ldots,m+n+1$, are c_j and the components of $\underline{g}_j$, respectively, where j is any integer from $[1,m]$. Specifically, by substituting the matrix (3.3) into equation (3.4), we find that $\ell_j(\underline{x}^+)$ is the sum

$$H_{m+1\,j} + \sum_{i=1}^{n} H_{m+i+1\,j}\,(\underline{x}^+ - \underline{x}_0)_i + \tfrac{1}{2}\sum_{i=1}^{m} H_{ij}\{(\underline{x}_i-\underline{x}_0)^T(\underline{x}^+-\underline{x}_0)\}^2, \qquad (4.17)$$

which is analogous to the form (3.7). Hence, because of the choice (4.9) of the components of $\underline{w}$, the symmetry of H gives the required result

$$\ell_j(\underline{x}^+) = \sum_{i=1}^{m+n+1} H_{ij}\,w_i = \underline{e}_j^T H\underline{w}, \qquad j=1,2,\ldots,m. \qquad (4.18)$$

In particular, the value (4.16) is just $\ell_t(\underline{x}^+)$. Moreover, some cancellation occurs if we combine expressions (4.11) and (4.15). These remarks and equation (4.14) imply that the parameters of the updating formula (4.12) take the values

$$\left.\begin{array}{ll}\alpha_t = \underline{e}_t^T H\underline{e}_t = H_{tt}, & \beta_t = \tfrac{1}{2}\|\underline{x}^+-\underline{x}_0\|^4 - \underline{w}^T H\underline{w},\\ \tau_t = \ell_t(\underline{x}^+), & \text{and} \quad \sigma_t = \alpha_t\,\beta_t + \tau_t^2.\end{array}\right\} \qquad (4.19)$$

The results (4.19) are not only useful in practice, but also they are relevant to the nearness of the matrix $W^+=(H^+)^{-1}$ to singularity. Indeed, formula (4.12) suggests that difficulties may arise from large elements of H^+ if $|\sigma_t|$ is unusually small. Further, we recall from Section 3 that we avoid singularity in W^+ by choosing t so that $\ell_t(\underline{x}^+)=\tau_t$ is nonzero. It follows from $\sigma_t=\alpha_t\,\beta_t+\tau_t^2$ that a nonnegative product $\alpha_t\,\beta_t$ would be welcome. Fortunately, we can establish the properties $\alpha_t\geq 0$ and $\beta_t \geq 0$ in theory, but the proof is given later, because it includes a convenient choice of $\underline{x}_0$, and the effects on H of changes to $\underline{x}_0$ are the subject of the next section.

5. Changes to the vector $\underline{x}_0$

As mentioned at the end of Section 1, the choice of $\underline{x}_0$ is important to the accuracy that is achieved in practice by the given Frobenius norm updating method and its applications. In particular, if $\underline{x}_0$ is unsuitable, and if the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$, are close to each other, which tends to happen towards the end of an unconstrained minimization calculation, then much cancellation occurs if $\ell_j(\underline{x}^+)$ is generated by formulae (3.6) and (3.7). This remark is explained, after the following fundamental property of $H=W^{-1}$ is established, where W is still the matrix

$$W \;=\; \left(\begin{array}{c|c} A & \underline{e} \;\vdots\; X^T \\ \hline \underline{e}^T \\ \cdots \\ X & 0 \end{array}\right). \tag{5.1}$$

Lemma 1: The leading $m\times m$ submatrix of $H=W^{-1}$ is independent of $\underline{x}_0$.

Proof: Let j be any integer from $[1,m]$. The definition of the Lagrange function $\ell_j(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, stated at the beginning of Section 3, does not depend on $\underline{x}_0$. Therefore the second derivative matrix (3.3) has this property too. Moreover, because the vector with the components H_{ij}, $i=1,2,\ldots,m+n+1$, is the j-th column of $H=W^{-1}$, it is orthogonal to the last $n+1$ columns of the matrix (5.1), which provides the conditions

$$\sum_{i=1}^m H_{ij} \;=\; 0 \qquad \text{and} \qquad \sum_{i=1}^m H_{ij}\,(\underline{x}_i-\underline{x}_0) \;=\; \sum_{i=1}^m H_{ij}\,\underline{x}_i \;=\; 0. \tag{5.2}$$

Thus the explicit occurrences of $\underline{x}_0$ on the right hand side of expression (3.3) can be removed, confirming that the matrix

$$\nabla^2\ell_j \;=\; \sum_{i=1}^m H_{ij}\,(\underline{x}_i-\underline{x}_0)\,(\underline{x}_i-\underline{x}_0)^T \;=\; \sum_{i=1}^m H_{ij}\,\underline{x}_i\,\underline{x}_i^T \tag{5.3}$$

is independent of $\underline{x}_0$. Therefore it is sufficient to prove that the elements H_{ij}, $i=1,2,\ldots,m$, can be deduced uniquely from the parts of equations (5.2) and (5.3) that are without $\underline{x}_0$.

We establish the equivalent assertion that, if the numbers μ_i, $i=1,2,\ldots,m$, satisfy the constraints

$$\left.\begin{array}{l} \sum_{i=1}^m \mu_i \;=\; 0, \qquad \sum_{i=1}^m \mu_i\,(\underline{x}_i-\underline{x}_0) \;=\; \sum_{i=1}^m \mu_i\,\underline{x}_i \;=\; 0 \\ \text{and} \quad \sum_{i=1}^m \mu_i\,(\underline{x}_i-\underline{x}_0)\,(\underline{x}_i-\underline{x}_0)^T \;=\; \sum_{i=1}^m \mu_i\,\underline{x}_i\,\underline{x}_i^T \;=\; 0 \end{array}\right\}, \tag{5.4}$$

then they are all zero. Let these conditions hold, and let the components of the vector $\underline{\mu}\in\mathcal{R}^{m+n+1}$ be μ_i, $i=1,2,\ldots,m$, followed by $n+1$ zeros. Because the submatrix A of the matrix (5.1) has the elements (2.7), the first m components of the product $W\underline{\mu}$ are the sums

$$\begin{aligned}(W\underline{\mu})_k &= \tfrac{1}{2}\sum_{i=1}^m \{(\underline{x}_k-\underline{x}_0)^T(\underline{x}_i-\underline{x}_0)\}^2\mu_i \\ &= \tfrac{1}{2}\,(\underline{x}_k-\underline{x}_0)^T\Big\{\sum_{i=1}^m \mu_i\,(\underline{x}_i-\underline{x}_0)\,(\underline{x}_i-\underline{x}_0)^T\Big\}\,(\underline{x}_k-\underline{x}_0) \\ &= 0, \qquad k=1,2,\ldots,m, \end{aligned} \tag{5.5}$$

the last equality being due to the second line of expression (5.4). Moreover, the definition (5.1) and the first line of expression (5.4) imply that the last $n+1$ components of $W\underline{\mu}$ are also zero. Hence the nonsingularity of W provides $\underline{\mu}=0$, which gives the required result. $\Box$

We now expose the cancellation that occurs in formulae (3.6) and (3.7) if all of the distances $\|\underline{x}^+-\underline{x}_b\|$ and $\|\underline{x}_i-\underline{x}_b\|$, $i=1,2,\ldots,m$, are bounded by 10ρ, say, but the number M, defined by $\|\underline{x}_0-\underline{x}_b\| = M\rho$, is large, $\underline{x}_b$ and ρ being taken from the trust region subproblem (1.4). We assume that the positions of the interpolation points give the property that the values $|\ell_j(\underline{x}^+)|$, $j=1,2,\ldots,m$, are not much greater than one. On the other hand, because of the Lagrange conditions (3.1) with $m\geq n+2$, some of the Lagrange functions have substantial curvature. Specifically, the magnitudes of some of the second derivative terms

$$\tfrac{1}{2}(\underline{x}_i-\underline{x}_b)^T\,\nabla^2\ell_j\,(\underline{x}_i-\underline{x}_b), \qquad 1\leq i,j\leq m, \tag{5.6}$$

are at least one, so some of the norms $\|\nabla^2\ell_j\|$, $j=1,2,\ldots,m$, are at least of magnitude ρ^{-2}. We consider the form (3.3) of $\nabla^2\ell_j$, after replacing $\underline{x}_0$ by $\underline{x}_b$, which is allowed by the conditions (5.2). It follows that some of the elements H_{kj}, $1\leq j,k\leq m$, are at least of magnitude ρ^{-4}, the integer m being a constant. Moreover, the positions of $\underline{x}_0$, $\underline{x}^+$ and $\underline{x}_i$, $i=1,2,\ldots,m$, imply that every scalar product (3.6) is approximately $M^2\rho^2$. Thus in practice formula (3.7) would include errors of magnitude M^4 times the relative precision of the computer arithmetic. Therefore the replacement of $\underline{x}_0$ by the current value of $\underline{x}_b$ is recommended if the ratio $\|\underline{x}_0-\underline{x}_b\|/\rho$ becomes large.

The reader may have noticed an easy way of avoiding the possible loss of accuracy that has just been mentioned. It is to replace $\underline{x}_0$ by $\underline{x}_b$ in formula (3.6), because then equation (3.7) remains valid without a factor of M^4 in the magnitudes of the terms under the summation sign. We have to retain $\underline{x}_0$ in the first line of expression (4.9), however, because formula (4.12) requires all components of the product $H\underline{w}$. Therefore a change to $\underline{x}_0$, as recommended at the end of the previous paragraph, can reduce some essential terms of the updating method by a factor of about M^4. We address the updating of H when $\underline{x}_0$ is shifted to $\underline{x}_0+\underline{s}$, say, but no modifications are made to the positions of the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$.

This task, unfortunately, requires $\mathcal{O}(n^3)$ operations in the case $m=\mathcal{O}(n)$ that is being assumed. Nevertheless, updating has some advantages over the direct calculation of $H = W^{-1}$ from the new W, one of them being stated in Lemma 1. The following description of a suitable procedure employs the vectors

$$\left.\begin{aligned} \underline{y}_k &= \underline{x}_k - \underline{x}_0 - \tfrac{1}{2}\,\underline{s} \\ \underline{z}_k &= (\underline{s}^T \underline{y}_k)\,\underline{y}_k + \tfrac{1}{4}\,\|\underline{s}\|^2\,\underline{s} \end{aligned}\right\}, \qquad k=1,2,\ldots,m, \tag{5.7}$$

because they provide convenient expressions for the changes to the elements of A. Specifically, the definitions (2.7) and (5.7) imply the identity

$$\begin{aligned} A_{ik}^{\text{new}} - A_{ik}^{\text{old}} &= \tfrac{1}{2}\,\{(\underline{x}_i - \underline{x}_0 - \underline{s})^T(\underline{x}_k - \underline{x}_0 - \underline{s})\}^2 - \tfrac{1}{2}\,\{(\underline{x}_i - \underline{x}_0)^T(\underline{x}_k - \underline{x}_0)\}^2 \\ &= \tfrac{1}{2}\,\{(\underline{y}_i - \tfrac{1}{2}\underline{s})^T(\underline{y}_k - \tfrac{1}{2}\underline{s})\}^2 - \tfrac{1}{2}\,\{(\underline{y}_i + \tfrac{1}{2}\underline{s})^T(\underline{y}_k + \tfrac{1}{2}\underline{s})\}^2 \\ &= \tfrac{1}{2}\,\{-\underline{s}^T\underline{y}_k - \underline{s}^T\underline{y}_i\}\,\{2\,\underline{y}_i^T\underline{y}_k + \tfrac{1}{2}\,\|\underline{s}\|^2\} \\ &= -\underline{z}_k^T\,\underline{y}_i - \underline{z}_i^T\,\underline{y}_k, \qquad 1 \le i,k \le m. \end{aligned} \tag{5.8}$$

Let Ω_X and Ω_A be the $(m+n+1)\times(m+n+1)$ matrices

$$\Omega_X = \left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & -\tfrac{1}{2}\underline{s} & I \end{array}\right) \quad \text{and} \quad \Omega_A = \left(\begin{array}{c|c|c} I & 0 & -Z^T \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & I \end{array}\right), \tag{5.9}$$

where Z is the $n\times m$ matrix that has the columns $\underline{z}_k$, $k=1,2,\ldots,m$. We find in the next paragraph that W can be updated by applying the formula

$$W^{\text{new}} = \Omega_X\,\Omega_A\,\Omega_X\,W^{\text{old}}\,\Omega_X^T\,\Omega_A^T\,\Omega_X^T. \tag{5.10}$$

The matrix Ω_X has the property that the product $\Omega_X W^{\text{old}}$ can be formed by subtracting $\frac{1}{2}s_i\,\underline{e}^T$ from the i-th row of X in expression (5.1) for $i=1,2,\ldots,n$. Thus X is overwritten by the $n\times m$ matrix Y, say, that has the columns $\underline{y}_k$, $k=1,2,\ldots,m$, defined by equation (5.7). Moreover, Ω_A is such that the pre-multiplication of $\Omega_X W^{\text{old}}$ by Ω_A changes only the first m rows of the current matrix, the scalar product of $\underline{z}_i$ with the k-th column of Y being subtracted from the k-th element of the i-th row of A^{old} for $i=1,2,\ldots,m$ and $k=1,2,\ldots,m$, which gives the $-\underline{z}_i^T\underline{y}_k$ term of the change from A^{old} to A^{new}, shown in the identity (5.8). Similarly, the post-multiplication of $\Omega_A\Omega_X W^{\text{old}}$ by Ω_X^T causes Y^T to occupy the position of X^T in expression (5.1), and then post-multiplication by Ω_A^T provides the other term of the identity (5.8), so A^{new} is the leading $m\times m$ submatrix of $\Omega_A\Omega_X W^{\text{old}}\Omega_X^T\Omega_A^T$. Finally, the outermost products of formula (5.10) overwrite Y and Y^T by the new X and the new X^T, respectively, which completes the updating of W.

The required new matrix H is the inverse of W^{new}. Therefore equation (5.10) implies the formula

$$H^{\text{new}} = (\Omega_X^T)^{-1}\,(\Omega_A^T)^{-1}\,(\Omega_X^T)^{-1}\,H^{\text{old}}\,\Omega_X^{-1}\,\Omega_A^{-1}\,\Omega_X^{-1}. \tag{5.11}$$

Moreover, the definitions (5.9) imply that the transpose matrices Ω_X^T and Ω_A^T have the inverses

$$(\Omega_X^T)^{-1} = \left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline 0 & 1 & \frac{1}{2}\underline{s}^T \\ \hline 0 & 0 & I \end{array}\right) \quad \text{and} \quad (\Omega_A^T)^{-1} = \left(\begin{array}{c|c|c} I & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline Z & 0 & I \end{array}\right), \qquad (5.12)$$

Expressions (5.11) and (5.12) provide a way of calculating H^{new} from H^{old} that is analogous to the method of the previous paragraph. Specifically, it is as follows.

The pre-multiplication of a matrix by $(\Omega_X^T)^{-1}$ is done by adding $\frac{1}{2}s_i$ times the $(m+i+1)$-th row of the matrix to the $(m+1)$-th row for $i=1,2,\ldots,n$, and the post-multiplication of a matrix by Ω_X^{-1} adds $\frac{1}{2}s_i$ times the $(m+i+1)$-th column of the matrix to the $(m+1)$-th column for the same values of i. Thus the symmetric matrix $(\Omega_X^T)^{-1}H^{\text{old}}\Omega_X^{-1} = H^{\text{int}}$, say, is calculated, and its elements differ from those of H^{old} only in the $(m+1)$-th row and column. Then the pre-multiplication of H^{int} by $(\Omega_A^T)^{-1}$ adds $(\underline{z}_k)_i$ times the k-th row of H^{int} to the $(m+i+1)$-th row of H^{int} for $k=1,2,\ldots,m$ and $i=1,2,\ldots,n$. This description also holds for post-multiplication of a matrix by Ω_A^{-1} if the two occurrences of "row" are replaced by "column". These operations yield the symmetric matrix $(\Omega_A^T)^{-1}H^{\text{int}}\Omega_A^{-1} = H^{\text{next}}$, say, so the elements of H^{next} are different from those of H^{int} only in the last n rows and columns. Finally, H^{new} is constructed by forming the product $(\Omega_X^T)^{-1}H^{\text{next}}\Omega_X^{-1}$ in the way that is given above. One feature of this procedure is that the leading $m\times m$ submatrices of H^{old}, H^{int}, H^{next} and H^{new} are all the same, which provides another proof of Lemma 1.

All the parameters (4.19) of the updating formula (4.12) are also independent of $\underline{x}_0$ in exact arithmetic. The definition $\alpha_t = H_{tt}$ and Lemma 1 imply that α_t has this property. Moreover, because the Lagrange function $\ell_t(\underline{x})$, $\underline{x}\in\mathcal{R}^n$, does not depend on $\underline{x}_0$, as mentioned at the beginning of the proof of Lemma 1, the parameter $\tau_t = \ell_t(\underline{x}^+)$ has this property too. We see in expression (4.19) that β_t is independent of t, and its independence of $\underline{x}_0$ is shown in the proof below of the last remark of Section 4. It follows that $\sigma_t = \alpha_t\beta_t + \tau_t^2$ is also independent of $\underline{x}_0$.

Lemma 2: Let H be the inverse of the matrix (5.1) and let $\underline{w}$ have the components (4.9). Then the parameters α_t and β_t of the updating formula (4.12) are nonnegative.

Proof: We write H in the partitioned form

$$H = W^{-1} = \left(\begin{array}{c|c} A & B^T \\ \hline B & 0 \end{array}\right)^{-1} = \left(\begin{array}{c|c} V & U^T \\ \hline U & \Upsilon \end{array}\right), \qquad (5.13)$$

where B is the bottom left submatrix of expression (5.1), and where the size of V is $m\times m$. Moreover, we recall from condition (2.10) that A has no negative eigenvalues. Therefore V and Υ are without negative and positive eigenvalues, respectively, which is well known and which can be shown as follows. Expression (5.13) gives the equations $VA+U^TB=I$ and $BV=0$, which imply the identity

$$\underline{\nu}^T V \underline{\nu} \;=\; \underline{\nu}^T (VA+U^TB)\, V \underline{\nu} \;=\; (V\underline{\nu})^T A\, (V\underline{\nu}), \qquad \underline{\nu} \in \mathcal{R}^m. \tag{5.14}$$

Thus the positive semidefiniteness of V is inherited from A. Expression (5.13) also gives $AU^T + B^T\Upsilon = 0$ and $UB^T = I$, which provide the equation

$$0 \;=\; U(AU^T + B^T\Upsilon) \;=\; UAU^T + \Upsilon, \tag{5.15}$$

so the negative semidefiniteness of Υ is also inherited from A.

By combining the positive semidefiniteness of V with formulae (4.19) and (5.13), we obtain the first of the required results

$$\alpha_t \;=\; H_{tt} \;=\; V_{tt} \;\geq\; 0. \tag{5.16}$$

Furthermore, we consider the value (4.19) of β_t in the special case $\underline{x}_0 = \underline{x}^+$. Then the term $\|\underline{x}^+ - \underline{x}_0\|$ is zero, and the definition (4.9) reduces to $\underline{w} = \underline{e}_{m+1}$. Thus, by using equation (5.13) and the negative semidefiniteness of Υ, we deduce that the value (4.19) of β_t achieves the required condition

$$\beta_t \;=\; -\underline{e}_{m+1}^T H\, \underline{e}_{m+1} \;=\; -H_{m+1\, m+1} \;=\; -\Upsilon_{11} \;\geq\; 0. \tag{5.17}$$

Of course this argument is not valid for other choices of $\underline{x}_0$. Fortunately, however, the conclusion $\beta_t \geq 0$ is preserved if any change is made to $\underline{x}_0$, because we find below that β_t is independent of $\underline{x}_0$.

If $\alpha_t = H_{tt}$ is zero, then, because V is positive semidefinite in equation (5.13), all the elements H_{it}, $i=1,2,\ldots,m$, are zero. It follows from equation (5.3) that the Lagrange function $\ell_t(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, is a linear polynomial. If this case occurs, then we make a tiny change to the positions of the interpolation points so that $\nabla^2 \ell_t$ becomes nonzero. The resultant change to β_t can be made arbitrarily small, because W is nonsingular. Therefore it is sufficient to prove that β_t is independent of $\underline{x}_0$ in the case $\alpha_t > 0$.

We deduce from equations (4.12), (4.14) and (4.16) that the t-th diagonal element of H^+ has the value

$$\begin{aligned} H_{tt}^+ \;&=\; \underline{e}_t^T H^+ \underline{e}_t \;=\; \alpha_t + \sigma_t^{-1} \left[\alpha_t (1-\tau_t)^2 - \beta_t \alpha_t^2 + 2\, \tau_t\, \alpha_t\, (1-\tau_t) \right] \\ &=\; \alpha_t + \sigma_t^{-1} \left[\alpha_t - \alpha_t \tau_t^2 - \beta_t \alpha_t^2 \right] \;=\; \alpha_t \,/\, (\alpha_t \beta_t + \tau_t^2). \end{aligned} \tag{5.18}$$

Now we have noted already that $\alpha_t = H_{tt}$ and $\tau_t = \ell_t(\underline{x}^+)$ are independent of $\underline{x}_0$, and Lemma 1 can be applied to the new matrix H^+, which shows that H_{tt}^+ is also independent of $\underline{x}_0$. It follows from equation (5.18) that β_t is independent of $\underline{x}_0$ when α_t is positive, which completes the proof. $\square$

6. Lagrange functions without their constant terms

We recall two reasons for calculating the values $\ell_j(\underline{x}^+)$, $j=1,2,\ldots,m$, of the Lagrange functions. One is that the integer t of expression (3.5) is chosen so that $\ell_t(\underline{x}^+)$ is nonzero, because then W^+ inherits nonsingularity from W. Secondly, equation (4.18) shows that these values are the first m components of $H\underline{w}$ in the updating formula (4.12). We find in this section that it may be advantageous to express $\ell_j(\underline{x}^+)$ as the sum

$$\ell_j(\underline{x}^+) \;=\; \{\ell_j(\underline{x}^+) - \ell_j(\underline{x}_b)\} + \delta_{jb}, \qquad j=1,2,\ldots,m, \tag{6.1}$$

where b is still the integer in $[1,m]$ such that $\underline{x}_b$ is the best of the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$. We deduce from equations (3.3) and (3.4) that the term $\ell_j(\underline{x}^+)-\ell_j(\underline{x}_b)$ has the value

$$\begin{aligned} &\underline{g}_j^T(\underline{x}^+-\underline{x}_b) + \tfrac{1}{2}\sum_{k=1}^{m} H_{kj}\left[\{(\underline{x}_k-\underline{x}_0)^T(\underline{x}^+-\underline{x}_0)\}^2 - \{(\underline{x}_k-\underline{x}_0)^T(\underline{x}_b-\underline{x}_0)\}^2\right] \\ &\qquad = \underline{g}_j^T\underline{d} + \sum_{k=1}^{m} H_{kj}\,\{(\underline{x}_k-\underline{x}_0)^T\underline{d}\}\,\{(\underline{x}_k-\underline{x}_0)^T(\underline{x}^{\rm mid}-\underline{x}_0)\}, \end{aligned} \tag{6.2}$$

$\underline{d}$ and $\underline{x}^{\rm mid}$ being the vectors

$$\underline{d} \;=\; \underline{x}^+-\underline{x}_b \qquad \text{and} \qquad \underline{x}^{\rm mid} \;=\; \tfrac{1}{2}\,(\underline{x}_b+\underline{x}^+). \tag{6.3}$$

Further, because the components of $\underline{g}_j$ are H_{ij}, $i=m+2,m+3,\ldots,m+n+1$, these remarks give the formula

$$\ell_j(\underline{x}^+) \;=\; \underline{e}_j^T H\,\widehat{\underline{w}} + \delta_{jb}, \qquad j=1,2,\ldots,m, \tag{6.4}$$

where $\widehat{\underline{w}}$ is the vector in $\mathcal{R}^{m+n+1}$ with the components

$$\left.\begin{aligned} \widehat{w}_k &= \tfrac{1}{2}\,[\,\{(\underline{x}_k-\underline{x}_0)^T(\underline{x}^+-\underline{x}_0)\}^2 - \{(\underline{x}_k-\underline{x}_0)^T(\underline{x}_b-\underline{x}_0)\}^2\,] \\ &= \{(\underline{x}_k-\underline{x}_0)^T\underline{d}\}\,\{(\underline{x}_k-\underline{x}_0)^T(\underline{x}^{\rm mid}-\underline{x}_0)\}, \quad k=1,2,\ldots,m, \\ \widehat{w}_{m+1} &= 0, \qquad \text{and} \qquad \widehat{w}_{m+i+1} \;=\; d_i, \qquad i=1,2,\ldots,n. \end{aligned}\right\} \tag{6.5}$$

Equation (6.4) has the advantage over expression (3.7) of tending to give better accuracy in practice when $\|\underline{d}\|=\|\underline{x}^+-\underline{x}_b\|$ is much smaller than $\|\underline{x}^+-\underline{x}_0\|$. Indeed, if $\|\underline{d}\|$ tends to zero, then equation (6.4) provides $\ell_j(\underline{x}^+)\rightarrow\delta_{jb}$ automatically in floating point arithmetic. Expression (3.7), however, includes the constant term $c_j=\ell_j(\underline{x}_0)$, which is typically of magnitude $\|\underline{x}_0-\underline{x}_b\|^2/\rho^2$, in the case when the distances $\|\underline{x}_i-\underline{x}_b\|$, $i=1,2,\ldots,m$, are not much greater than the trust region radius ρ. Thus, if $\|\underline{d}\|$ tends to zero, then the contributions to formula (3.7) from the errors in c_j, $j=1,2,\ldots,m$, become relatively large.

Another advantage of equation (6.4), which provides the challenge that is addressed in the remainder of this section, is that the $(m+1)$-th column of H is not required, because $\widehat{w}_{m+1}$ is zero. Therefore we let Θ be the $(m+n)\times(m+n)$ symmetric matrix that is formed by suppressing the $(m+1)$-th row and column of H, and we seek convenient versions of the calculations that have been described already, when Θ is stored instead of H. In particular, the new version of equation (6.4) is the formula

$$\ell_j(\underline{x}^+) \;=\; \underline{e}_j^T \Theta\, \underline{\check{w}} + \delta_{jb}, \qquad j=1,2,\ldots,m, \tag{6.6}$$

where $\underline{e}_j$ is now the j-th coordinate vector in $\mathcal{R}^{m+n}$, and where $\underline{\check{w}}$ is $\underline{\widehat{w}}$ without its $(m+1)$-th component.

The modifications to the work of Section 5 are straightforward. Indeed, the pre-multiplications by $(\Omega_X^T)^{-1}$ and the post-multiplications by Ω_X^{-1} in expression (5.11) change only the $(m+1)$-th row and column, respectively, of the current matrix. Therefore they are irrelevant to the calculation of Θ^{new} from Θ^{old}, say, which revises Θ when $\underline{x}_0$ is replaced by $\underline{x}_0+\underline{s}$. Further, pre-multiplication of an $(m+n+1)\times(m+n+1)$ matrix by $(\Omega_A^T)^{-1}$ adds linear combinations of the first m rows to the last n rows, and post-multiplication by Ω_A^{-1} operates similarly on the columns instead of the rows of the current matrix. It follows from equations (5.11) and (5.12) that Θ^{new} is the product

$$\Theta^{\text{new}} \;=\; \Xi\, \Theta^{\text{old}}\, \Xi^T, \tag{6.7}$$

where Ξ is the $(n+m)\times(n+m)$ matrix

$$\Xi \;=\; \left(\begin{array}{c|c} I & 0 \\ \hline Z & I \end{array}\right), \tag{6.8}$$

which is constructed by deleting the $(m+1)$-th row and column of $(\Omega_A^T)^{-1}$ in expression (5.12). The fact that formula (6.7) requires less computation than formula (5.11) is welcome. Specifically, the multiplications by Ξ and Ξ^T are done by forming the linear combinations of rows and columns of the current matrix that are implied by the definition (6.8).

Our development of a suitable way of updating Θ, when the change (3.5) is made to the interpolation points, depends on the identity

$$\underline{\widehat{w}} \;=\; \underline{w} - W\underline{e}_b. \tag{6.9}$$

It is a consequence of equation (4.9) and the definition (6.5) of $\underline{\widehat{w}}$, because $W\underline{e}_b$ is the b-th column of the matrix (5.1). By multiplying this identity by $H=W^{-1}$, we find the relation

$$\underline{e}_t - H\,\underline{w} \;=\; \underline{e}_t - \underline{e}_b - H\,\underline{\widehat{w}}, \tag{6.10}$$

which is useful, because the value $\widehat{w}_{m+1}=0$ implies that the first m and last n components of $H\,\underline{\widehat{w}}$ are the components of the product $\Theta\,\underline{\check{w}}$ of equation (6.6). Moreover, for every integer t in $[1,m]$, the first m and last n components of $H\,\underline{e}_t$ are the components of $\Theta\,\underline{e}_t$, the coordinate vectors $\underline{e}_t$ being in $\mathcal{R}^{m+n+1}$ and in $\mathcal{R}^{m+n}$, respectively. We let Θ^+ be the $(m+n)\times(m+n)$ symmetric matrix that is formed by suppressing the $(m+1)$-th row and column of H^+, and we suppress these parts of the right hand side of the updating formula (4.12), after making the substitution (6.10). It follows that Θ^+ is the matrix

$$\begin{aligned}\Theta^+ \;&=\; \Theta + \frac{1}{\sigma_t}\Big[\alpha_t\,(\underline{e}_t-\underline{e}_b-\Theta\,\underline{\check{w}})\,(\underline{e}_t-\underline{e}_b-\Theta\,\underline{\check{w}})^T - \beta_t\,\Theta\,\underline{e}_t\,\underline{e}_t^T\Theta \\ &\qquad + \tau_t\Big\{\Theta\,\underline{e}_t\,(\underline{e}_t-\underline{e}_b-\Theta\,\underline{\check{w}})^T + (\underline{e}_t-\underline{e}_b-\Theta\,\underline{\check{w}})\,\underline{e}_t^T\Theta\Big\}\Big], \qquad (6.11)\end{aligned}$$

the parameters α_t, β_t, τ_t and σ_t being as before, but all the coordinate vectors are now in $\mathcal{R}^{m+n}$.

Expression (4.19) shows that the values of the parameters $\alpha_t=H_{tt}=\Theta_{tt}$ and $\tau_t=\ell_t(\underline{x}^+)$ are available, and also we apply $\sigma_t=\alpha_t\,\beta_t+\tau_t^2$ after calculating β_t in a new way. Specifically, equations (4.19) and (6.9), with the definitions of Θ and $\underline{\check{w}}$, provide the form

$$\begin{aligned}\beta_t \;&=\; \tfrac{1}{2}\,\|\underline{x}^+-\underline{x}_0\|^4 - (\underline{\widehat{w}}+W\,\underline{e}_b)^T H\,(\underline{\widehat{w}}+W\,\underline{e}_b) \\ &=\; \tfrac{1}{2}\,\|\underline{x}^+-\underline{x}_0\|^4 - \underline{\check{w}}^T\Theta\,\underline{\check{w}} - 2\,\underline{e}_b^T\,\underline{\widehat{w}} - W_{bb}, \qquad (6.12)\end{aligned}$$

because $\widehat{w}_{t+1}$ is zero and W is the inverse of H. Further, we find in the definition (6.5) that $2\,\underline{e}_b^T\underline{\widehat{w}}$ is the difference

$$2\,\underline{e}_b^T\,\underline{\widehat{w}} \;=\; 2\,\widehat{w}_b \;=\; \{(\underline{x}_b-\underline{x}_0)^T(\underline{x}^+-\underline{x}_0)\}^2 - \|\underline{x}_b-\underline{x}_0\|^4, \qquad (6.13)$$

and equation (2.7) gives $W_{bb}=A_{bb}=\frac{1}{2}\,\|\underline{x}_b-\underline{x}_0\|^4$. Therefore β_t has the value

$$\begin{aligned}\beta_t \;&=\; \tfrac{1}{2}\,\|\underline{x}^+-\underline{x}_0\|^4 - \{(\underline{x}_b-\underline{x}_0)^T(\underline{x}^+-\underline{x}_0)\}^2 + \tfrac{1}{2}\,\|\underline{x}_b-\underline{x}_0\|^4 - \underline{\check{w}}^T\Theta\,\underline{\check{w}} \\ &=\; \{(\underline{x}^{\rm mid}-\underline{x}_0)^T\underline{d}\}^2 + \|\underline{x}^{\rm mid}-\underline{x}_0\|^2\,\|\underline{d}\|^2 - \underline{\check{w}}^T\Theta\,\underline{\check{w}}, \qquad (6.14)\end{aligned}$$

where the last line is derived by expressing $\underline{x}^+$ and $\underline{x}_b$ in terms of the vectors (6.3). Thus the calculation of β_t is also straightforward, which completes the description of the updating of Θ when an interpolation point is moved. The amount of work of this method is about the same as the effort of updating H by formula (4.12). Some numerical experiments on the stability of long sequences of updates of both H and Θ are reported in Section 7.

In one of those experiments, namely Test 5, substantial errors are introduced into the initial matrix Θ deliberately. Then, after a sequence of updates that moves all the interpolation points from their initial positions, the form (6.6) of the Lagrange functions, where $\underline{x}^+$ is now a general point of $\mathcal{R}^n$, provides the Lagrange conditions $\ell_j(\underline{x}_i)=\delta_{ij}$, $1\le i,j\le m$, to high accuracy. It seems, therefore, that the updating method of this section enjoys a stability property that is similar to the one that is addressed in the second paragraph of Section 4. This conjecture is established below, most of the analysis being the proof of the following lemma, which may be skipped by the reader without loss of continuity. The lemma was suggested by numerical calculations of all the products $\Theta\Omega$ in a sequence of applications of formula (6.11), where Ω is the $(m{+}n)\times(m{+}n)$ matrix that is constructed by deleting the $(m+1)$-th row and column of W.

Lemma 3: Let the updating method of this section calculate Θ^+, where the symmetric matrix Θ and the interpolation points $\underline{x}_i$, $i=1,2,\ldots,m$, are such that the denominator σ_t of formula (6.11) is nonzero. Then the t-th and b-th columns of $\Theta^+\Omega^+{-}I$ are the same, where Ω^+ is the matrix Ω for the new positions (3.5) of the interpolation points. Further, if p is any integer in $[1,m]$ such that the p-th and b-th columns of $\Theta\,\Omega-I$ are the same, then this property is inherited by the p-th and b-th columns of $\Theta^+\Omega^+{-}I$.

Proof: We begin by assuming $t\neq b$, because otherwise the first statement of the lemma is trivial. Therefore we can write the first line of expression (6.14) in the form

$$\beta_t \;=\; (\underline{e}_t-\underline{e}_b)^T\Omega^+(\underline{e}_t-\underline{e}_b) - \underline{\check{w}}^T\Theta\,\underline{\check{w}}. \tag{6.15}$$

Moreover, the definition (6.5) shows that, with the exception of $\widehat{w}_t$, the components of $\underline{\widehat{w}}$ are those of $W^+(\underline{e}_t-\underline{e}_b)$. Hence the construction of Ω^+ and $\underline{\check{w}}$ from W^+ and $\underline{\widehat{w}}$ gives the identity

$$\Omega^+(\underline{e}_t-\underline{e}_b) \;=\; \underline{\check{w}}+\xi\,\underline{e}_t \tag{6.16}$$

for some $\xi\in\mathcal{R}$. We consider the vector $\Theta^+\Omega^+(\underline{e}_t-\underline{e}_b)$, where Θ^+ is the matrix (6.11), because the first assertion of the lemma is equivalent to the condition

$$\Theta^+\Omega^+(\underline{e}_t-\underline{e}_b) \;=\; \underline{e}_t-\underline{e}_b. \tag{6.17}$$

Equations (6.15) and (6.16) are useful as they provide the scalar products

$$\left.\begin{aligned}(\underline{e}_t-\underline{e}_b-\Theta\,\underline{\check{w}})^T\Omega^+(\underline{e}_t-\underline{e}_b) \;&=\; \beta_t-\xi\,\underline{\check{w}}^T\Theta\,\underline{e}_t \;=\; \beta_t-\xi\,\tau_t\\ \underline{e}_t^T\Theta\,\Omega^+(\underline{e}_t-\underline{e}_b) \;&=\; \underline{e}_t^T\Theta\,\underline{\check{w}}+\xi\,\underline{e}_t^T\Theta\,\underline{e}_t \;=\; \tau_t+\xi\,\alpha_t\end{aligned}\right\}, \tag{6.18}$$

the right hand sides being obtained from formulae (6.6) and (4.19). Thus expressions (6.11) and (6.16) with $\sigma_t=\alpha_t\beta_t+\tau_t^2$ imply the required result

$$\Theta^+\Omega^+(\underline{e}_t-\underline{e}_b)=\Theta\,(\underline{\check{w}}+\xi\,\underline{e}_t)+\frac{1}{\sigma_t}\Big[\sigma_t(\underline{e}_t-\underline{e}_b-\Theta\,\underline{\check{w}})-\xi\,\sigma_t\,\Theta\,\underline{e}_t\Big]=\underline{e}_t-\underline{e}_b. \tag{6.19}$$

In the remainder of the proof we assume $p\neq b$ and $p\neq t$, because the second assertion of the lemma is trivial in the case $p=b$, and the analysis of the previous paragraph applies in the case $p=t$. We also assume $t\neq b$ for the moment, and will address the alternative $t\!=\!b$ later. Therefore, because all differences between the matrices Ω and Ω^+ are confined to their t-th rows and columns, the equation

$$\Omega^+(\underline{e}_p\!-\!\underline{e}_b) \;=\; \Omega\,(\underline{e}_p\!-\!\underline{e}_b) + \eta\,\underline{e}_t \tag{6.20}$$

holds for some $\eta\!\in\!\mathcal{R}$. Further, expressions (6.16) and (6.20) provide the identity

$$(\underline{e}_t\!-\!\underline{e}_b\!-\!\Theta\,\underline{\breve{w}})^T\Omega^+(\underline{e}_p\!-\!\underline{e}_b) \;=\; (\underline{\breve{w}}\!+\!\xi\,\underline{e}_t)^T(\underline{e}_p\!-\!\underline{e}_b) - \underline{\breve{w}}^T\Theta\,\{\Omega\,(\underline{e}_p\!-\!\underline{e}_b) + \eta\,\underline{e}_t\}. \tag{6.21}$$

It follows from the hypothesis

$$(\Theta\,\Omega - I)\,(\underline{e}_p\!-\!\underline{e}_b) \;=\; 0 \tag{6.22}$$

and equation (6.20) that the scalar products of $\Theta^+\Omega^+(\underline{e}_p\!-\!\underline{e}_b)$ have the values

$$\left.\begin{aligned} (\underline{e}_t\!-\!\underline{e}_b\!-\!\Theta\,\underline{\breve{w}})^T\Omega^+(\underline{e}_p\!-\!\underline{e}_b) \;&=\; \xi\,\underline{e}_t^T(\underline{e}_p\!-\!\underline{e}_b) - \eta\,\underline{\breve{w}}^T\Theta\,\underline{e}_t \;=\; -\eta\,\tau_t \\ \underline{e}_t^T\Theta\,\Omega^+(\underline{e}_p\!-\!\underline{e}_b) \;&=\; \underline{e}_t^T(\underline{e}_p\!-\!\underline{e}_b\!+\!\eta\,\Theta\,\underline{e}_t) \;=\; \eta\,\alpha_t \end{aligned}\right\}. \tag{6.23}$$

Thus equations (6.11), (6.20) and (6.22) with $\sigma_t\!=\!\alpha_t\beta_t\!+\!\tau_t^2$ give the condition

$$\Theta^+\Omega^+(\underline{e}_p\!-\!\underline{e}_b) \;=\; \Theta\,\Omega\,(\underline{e}_p\!-\!\underline{e}_b) + \eta\,\Theta\,\underline{e}_t + \frac{1}{\sigma_t}\Big[-\eta\,\sigma_t\,\Theta\,\underline{e}_t\Big] \;=\; \underline{e}_p\!-\!\underline{e}_b, \tag{6.24}$$

which shows that the p-th and b-th columns of $\Theta^+\Omega^+\!-\!I$ are the same.

When $t\!=\!b$ and $p\!\neq\!b$, only the t-th component of $(\Omega^+\!-\!\Omega)\underline{e}_p$ can be nonzero, but the definition (6.5) shows that $(\Omega^+\!-\!\Omega)\underline{e}_b$ is the sum of $\underline{\breve{w}}$ and a multiple of $\underline{e}_t$. Therefore the analogue of equation (6.20) in the present case is the expression

$$\Omega^+(\underline{e}_p\!-\!\underline{e}_b) \;=\; \Omega\,(\underline{e}_p\!-\!\underline{e}_b) - \underline{\breve{w}} + \zeta\,\underline{e}_b \tag{6.25}$$

for some $\zeta\!\in\!\mathcal{R}$. We require a relation between ζ and β_t, so, by taking the scalar product of this expression with $\underline{e}_b$, we find the value

$$\zeta \;=\; \underline{e}_p^T(\Omega^+\!-\!\Omega)\,\underline{e}_b - \Omega_{bb}^+ + \Omega_{bb} + \underline{e}_b^T\underline{\breve{w}} \;=\; \underline{e}_p^T\underline{\breve{w}} - \Omega_{bb}^+ + \Omega_{bb} + \underline{e}_b^T\underline{\breve{w}}. \tag{6.26}$$

Moreover, we write the last line of equation (6.12) in the form

$$\beta_t \;=\; \Omega_{bb}^+ - \underline{\breve{w}}^T\Theta\,\underline{\breve{w}} - 2\,\underline{e}_b^T\underline{\breve{w}} - \Omega_{bb}. \tag{6.27}$$

Thus we obtain the identity

$$\beta_t \;=\; \underline{\breve{w}}^T(\underline{e}_p\!-\!\underline{e}_b) - \underline{\breve{w}}^T\Theta\,\underline{\breve{w}} - \zeta, \tag{6.28}$$

which is useful for simplifying one of the scalar products that occur when expressions (6.11) and (6.25) are substituted into $\Theta^+\Omega^+(\underline{e}_p\!-\!\underline{e}_b)$. Indeed, because formula (6.6) provides $\tau_t = \ell_t(\underline{x}^+) = \underline{e}_b^T \Theta\, \underline{\check{w}} + 1$, and because the hypothesis (6.22) still holds, the relation (6.25) gives the values

$$\left. \begin{array}{rcl} (-\Theta\, \underline{\check{w}})^T \Omega^+ (\underline{e}_p - \underline{e}_b) & = & -\underline{\check{w}}^T (\underline{e}_p - \underline{e}_b) + \underline{\check{w}}^T \Theta\, \underline{\check{w}} - \zeta\, \underline{\check{w}}^T \Theta\, \underline{e}_b \;=\; -\beta_t - \zeta\, \tau_t \\ \underline{e}_t^T \Theta\, \Omega^+ (\underline{e}_p - \underline{e}_b) & = & \underline{e}_b^T (\underline{e}_p - \underline{e}_b) - \underline{e}_b^T \Theta\, \underline{\check{w}} + \zeta\, \underline{e}_b^T \Theta\, \underline{e}_b \;=\; -\tau_t + \zeta\, \alpha_t \end{array} \right\} . \tag{6.29}$$

Further, $\underline{e}_t - \underline{e}_b$ is zero in formula (6.11). It follows from the equations (6.25), (6.22) and $\sigma_t = \alpha_t \beta_t + \tau_t^2$ that the required condition

$$\begin{array}{rcl} \Theta^+ \Omega^+ (\underline{e}_p - \underline{e}_b) & = & \Theta\, \Omega\, (\underline{e}_p - \underline{e}_b) - \Theta\, \underline{\check{w}} + \zeta\, \Theta\, \underline{e}_b + \dfrac{1}{\sigma_t} \left[\sigma_t\, \Theta\, \underline{\check{w}} - \zeta\, \sigma_t\, \Theta\, \underline{e}_t \right] \\ & = & \Theta\, \Omega\, (\underline{e}_p - \underline{e}_b) \;=\; \underline{e}_p - \underline{e}_b \end{array} \tag{6.30}$$

is achieved. The proof is complete. $\square$

The hypothesis (6.22) is important to practical calculations for the following reason. Let Θ be any symmetric matrix that satisfies this hypothesis for some integer $p \in [1, m]$, and, for the moment, let $\underline{x}^+$ be the interpolation point $\underline{x}_p$. Then expression (6.5) gives the vector $\underline{\widehat{w}} = W \underline{e}_p - W \underline{e}_b$, which implies the equation $\underline{\check{w}} = \Omega \underline{e}_p - \Omega \underline{e}_b$, because of the construction of $\underline{\check{w}}$ and Ω from $\underline{\widehat{w}}$ and W. It follows from condition (6.22) that the right hand side of formula (6.6) takes the value

$$\underline{e}_j^T \Theta\, (\Omega\, \underline{e}_p - \Omega\, \underline{e}_b) + \delta_{jb} \;=\; \delta_{jp}, \qquad j = 1, 2, \ldots, m. \tag{6.31}$$

In other words, formula (6.6) agrees with the Lagrange conditions $\ell_j(\underline{x}_p) = \delta_{jp}$, $j = 1, 2, \ldots, m$, when the p-th and b-th columns of $\Theta\Omega - I$ are the same.

Let $\mathcal{B}$ be the subset of $\{1, 2, \ldots, m\}$ that is composed of such integers p. In exact arithmetic $\mathcal{B}$ contains the indices of all the interpolation points throughout the sequence of updating calculations, but we assume that there are substantial errors in the initial symmetric matrix Θ. The lemma shows that, after the updating for the change (3.5) to the positions of the interpolation points, the new $\mathcal{B}$ includes all the elements of the old $\mathcal{B}$ and also the integer t. Some other part of the iteration, however, may alter b to a new value, b^+ say, which preserves $\mathcal{B}$ if b^+ is in $\mathcal{B}$. This condition holds after the change (3.5) if b^+ is set to t, which is usual in applications to unconstrained optimization. For example, we find in the third paragraph of Section 1 that b is altered only in the case $F(\underline{x}^+) < F(\underline{x}_b)$, and then $b^+ = t$ is selected, in order to provide the property (1.3) on the next iteration. Thus it follows from the lemma that the number of elements of $\mathcal{B}$ increases monotonically. Further, $\mathcal{B}$ becomes the set $\{1, 2, \ldots, m\}$ when all the interpolation points have been moved from their initial positions, which means that all the first m columns of $\Theta\Omega - I$ are the same. Therefore, because of the last remark of the previous paragraph, the Lagrange conditions $\ell_j(\underline{x}_i) = \delta_{ij}$, $1 \le i, j \le m$, are achieved, as mentioned before the statement of Lemma 3.

7. Numerical results and discussion

The values $n=50$ and $m=2n+1=101$ are employed throughout our experiments on the numerical accuracy of the updating formulae (4.12) and (6.11). Further, in every case the initial positions of the m interpolation points are $\underline{x}_1=0$, $\underline{x}_{2j}=\underline{e}_j$ and $\underline{x}_{2j+1}=-\underline{e}_j$, $j=1,2,\ldots,n$, where $\underline{e}_j$ is still the j-th coordinate vector in $\mathcal{R}^n$. We let the fixed vector $\underline{x}_0$ be a multiple of $\underline{e}$, which is now the vector in $\mathcal{R}^n$ whose components are all one. Thus it is straightforward to find expressions for the elements of the initial matrices W, H and Θ analytically. Each new interpolation point $\underline{x}^+$ is generated randomly from the distribution that is uniform in the ball $\{\underline{x} : \|\underline{x}\| \leq \rho\} \subset \mathcal{R}^n$, where the norm is Euclidean, and where ρ is fixed at one in most of the experiments, but we study too whether the stability of the updating formulae is impaired by reductions in ρ. The subscript b of the "best" interpolation point $\underline{x}_b$ is set to one initially. Then, after each change (3.5) to the interpolation points, the new value of b is $b^+=b$ or $b^+=t$ in the case $\|\underline{x}^+\| \geq \|\underline{x}_b\|$ or $\|\underline{x}^+\| < \|\underline{x}_b\|$, respectively. We consider three different procedures for choosing the point $\underline{x}_t$ that is removed from the set $\{\underline{x}_i : i=1,2,\ldots,m\}$, in order to make room for $\underline{x}^+$. They pick $\underline{x}_t=\underline{x}_b$ occasionally, and then the new $\underline{x}_b$ is always $\underline{x}^+$, so the initial choice $\underline{x}_b=\underline{x}_1=0$ is temporary. The term "iteration" denotes a selection of $\underline{x}^+$ and t and the updating of both H and Θ for the change (3.5) to the interpolation points.

The results of seven experiments are reported, and in the first five ρ is fixed at one. The purpose of Test 1 and Test 2 is to compare two ways of picking the integer t on each iteration, and in these cases $\underline{x}_0$ is zero. In Test 1, we let t be the least integer from $[1,m]$ that satisfies the equation

$$|\,\ell_t(\underline{x}^+)\,| \;=\; \max\{\,|\,\ell_i(\underline{x}^+)\,| : i=1,2,\ldots,m\}, \tag{7.1}$$

which gives $\ell_t(\underline{x}^+) \neq 0$. On the other hand, because the second paragraph of Section 4 states that errors from previous iterations in the t-th column of $\Delta = W-H^{-1}$ are removed by the updating of H, we prefer to discard older interpolation points. In Test 2, therefore, equation (7.1) is relaxed to the condition

$$|\,\ell_t(\underline{x}^+)\,| \;\geq\; \tfrac{1}{2}\,\max\{\,|\,\ell_i(\underline{x}^+)\,| : i=1,2,\ldots,m\}, \tag{7.2}$$

in order to give some freedom in the choice of t. This freedom is taken up by letting $\underline{x}_t$ be the point that was introduced first from among the points that satisfy inequality (7.2). Ties occur when two or more of the eligible points are survivors from the initialization procedure, and then each tie is broken by the larger value of $|\ell_t(\underline{x}^+)|$.

When we look at the first table of results later, we are going to find that both procedures of the previous paragraph seem to provide good accuracy. The theoretical reason for the latter method is attractive, however, so the values of t in Tests 3 to 5 are taken from Test 2. Tests 3 and 4 show some consequences of choosing $\underline{x}_0$ to be nonzero, this constant vector being set to $n^{-1/2}\underline{e}$ and $\underline{e}$, respectively. Thus Test 3 has the property $\|\underline{x}_0\|=1$.

Of course, if errors do not build up substantially, then one cannot discover directly whether an updating formula is able to correct an accumulation of errors automatically. Therefore the difference between Test 2 and Test 5 is that, in the latter case, artificial errors are introduced into the H and Θ matrices before the first iteration. Then the updating formulae are applied in the usual way. Specifically, after the initial H has been found as usual, Test 5 perturbs each element on the diagonal and in the lower triangular part of H, where every perturbation is a random number from the distribution that is uniform on $[-10^{-4}, 10^{-4}]$. Then the upper triangular part of H is defined by symmetry. Further, the initial Θ is formed by deleting the $(m{+}1)$-th row and column of the initial H.

The tables investigate whether the conditions $\ell_j(\underline{x}_k) = \delta_{jk}$, $1 \leq j,k \leq m$, are satisfied adequately in practice, after a sequence of applications of the updating formula (4.12) or (6.11), the Lagrange function $\ell_j(\underline{x}^+)$, $\underline{x}^+ \in \mathcal{R}^n$, being defined by equation (4.18) or (6.4), respectively. If the point $\underline{x}^+$ of expression (4.18) were the interpolation point $\underline{x}_k$, then the definition (4.9) would set $\underline{w}$ to the k-th column of W, namely $W\underline{e}_k$. Thus equation (4.18) takes the form

$$\ell_j(\underline{x}_k) \;=\; \underline{e}_j^T H W \underline{e}_k, \qquad 1 \leq j,k \leq m. \tag{7.3}$$

It follows that the leading $m \times m$ submatrix of $HW - I$ gives the errors in the Lagrange conditions. Furthermore, for $j = 1, 2, \ldots, m$, the constraints on the coefficients of $\ell_j(\underline{x})$, $\underline{x} \in \mathcal{R}^n$, are the last $(n{+}1)$ equations of the system (2.8) when $\underline{\lambda}$ has the components H_{ji}, $i = 1, 2, \ldots, m$, so the top right $m \times (n{+}1)$ submatrix of HW gives the errors in these constraints. Therefore the entries to the left and right of each solidus sign in Table 1 are values of the expressions

$$\left.\begin{array}{l} \log_{10} \max \{\, |\, (HW)_{jk} - \delta_{jk} \,| \; : \; 1 \leq j,k \leq m \,\} \qquad \text{and} \\ \log_{10} \max \{\, |\, (HW)_{jk} \,| \; : \; 1 \leq j \leq m, \; m{+}1 \leq k \leq m{+}n{+}1 \,\} \end{array} \right\}, \tag{7.4}$$

respectively, after the numbers of iterations that are stated at the head of the table. All the calculations were coded in Fortran and run on a Sun Ultra 10 workstation in double precision arithmetic.

Table 1 shows, as mentioned already, that Tests 1 and 2 on different ways of choosing t provide about the same accuracy. We see also that the largest error in a constraint grows by about the factor $10^{1.5} \approx 30$ as the iteration count increases from 10^2 to 10^5. On the other hand, the accuracy in the Lagrange conditions remains excellent, due to the stability property that is the subject of the second paragraph of Section 4. Moreover, a comparison of Tests 2 to 4 indicates the deterioration that can arise from a poor choice of the fixed vector $\underline{x}_0$, which will receive more attention in Table 3. The Test 5 results are a triumph for the stability of the updating formula, the substantial errors that are introduced initially into the Lagrange conditions being corrected automatically to full machine precision. A precaution in the computer code delays this procedure slightly, however, as explained below.

	After 10^2 iterations	After 10^3 iterations	After 10^4 iterations	After 10^5 iterations
Test 1	$-14.8\,/-14.5$	$-14.8\,/-14.1$	$-14.6\,/-13.6$	$-14.5\,/-13.0$
Test 2	$-14.8\,/-14.4$	$-14.8\,/-14.1$	$-14.7\,/-13.5$	$-14.8\,/-13.0$
Test 3	$-12.9\,/-13.1$	$-13.8\,/-13.9$	$-13.6\,/-13.4$	$-13.7\,/-13.0$
Test 4	$-7.4\,/-10.6$	$-7.3\,/-10.4$	$-7.1\,/-10.2$	$-7.2\,/-10.3$
Test 5	$-3.3\,/-2.8$	$-14.9\,/-2.9$	$-14.8\,/-2.9$	$-14.8\,/-2.8$

Table 1: The values (7.4) of logarithms of errors for $\rho=1$

The precaution responds to the remark that, if H is any matrix that is symmetric and nonsingular, then the relation (4.1) between $\Delta = W - H^{-1}$ and $\Delta^+ = W^+ - (H^+)^{-1}$ is valid even if σ is zero in formula (4.12), which is allowed in theory, because only $(H^+)^{-1}$ has to be well-defined. The random perturbations to H in Test 5 can cause the updating formula to fail because $|\sigma|$ is too small, however, although Lemma 2 states that the parameters α_t and β_t of the equation $\sigma_t = \alpha_t \beta_t + \tau_t^2$ are nonnegative in the case $H = W^{-1}$, and the selected value of t satisfies $\tau_t \neq 0$. Therefore the computer code employs the formula

$$\sigma_t \;=\; \max[\,0, \alpha_t\,]\; \max[\,0, \beta_t\,] + \tau_t^2. \tag{7.5}$$

Thus σ_t may be different from $\alpha_t \beta_t + \tau_t^2$, and then the property (4.1) would not be achieved in exact arithmetic. In particular, the t-th diagonal element of Δ^+ would not be reduced to zero by the current iteration.

Next we derive analogues of the expressions (7.4) for the updating formula (6.11). We recall that the $(m+n) \times (m+n)$ matrix Ω is constructed by deleting the $(m+1)$-th row and column of W, which gives the identities

$$\underline{e}_j^T H W \underline{e}_k \;=\; \underline{e}_j^T \Theta\, \Omega\, \underline{e}_k + H_{j\,m+1} W_{m+1\,k}, \qquad 1 \leq j, k \leq m, \tag{7.6}$$

the coordinate vectors on the left and right hand sides being in $\mathcal{R}^{m+n+1}$ and in $\mathcal{R}^{m+n}$, respectively. Now $H_{j\,m+1}$ is the constant term c_j of the Lagrange function (3.4), which is suppressed by the methods of Section 6, and the matrix (5.1) includes the elements $W_{m+1\,k} = 1$, $k = 1, 2, \ldots, m$. It follows from equations (7.3) and (7.6) that formula (6.4) gives the Lagrange conditions $\ell_j(\underline{x}_k) = \delta_{jk}$, $1 \leq j, k \leq m$, if and only if Θ has the property

$$\underline{e}_j^T \Theta\, \Omega\, \underline{e}_k + \widehat{c}_j \;=\; \delta_{jk}, \qquad 1 \leq j, k \leq m, \tag{7.7}$$

for some real numbers $\widehat{c}_j$, $j = 1, 2, \ldots, m$. Therefore we let the analogue of the first part of expression (7.4) be the quantity

$$\log_{10} \min_{\widehat{\underline{c}} \in \mathcal{R}^m} \max \{\, |\, (\Theta\, \Omega)_{jk} + \widehat{c}_j - \delta_{jk}\,| \;:\; 1 \leq j, k \leq m \,\}. \tag{7.8}$$

	After 10^2 iterations	After 10^3 iterations	After 10^4 iterations	After 10^5 iterations
Test 1	$-14.6\,/\!-\!14.9$	$-14.5\,/\!-\!14.6$	$-14.6\,/\!-\!14.2$	$-14.6\,/\!-\!13.7$
Test 2	$-14.7\,/\!-\!15.0$	$-14.8\,/\!-\!14.7$	$-14.9\,/\!-\!14.1$	$-14.9\,/\!-\!13.7$
Test 3	$-13.9\,/\!-\!14.3$	$-14.0\,/\!-\!14.4$	$-14.1\,/\!-\!14.0$	$-14.1\,/\!-\!13.5$
Test 4	$-9.6\,/\!-\!11.3$	$-9.3\,/\!-\!11.4$	$-9.6\,/\!-\!11.7$	$-9.6\,/\!-\!11.6$
Test 5	$-3.6\,/\!-\!3.4$	$-15.0\,/\!-\!3.5$	$-15.0\,/\!-\!3.4$	$-14.9\,/\!-\!3.5$

Table 2: The values (7.8) and (7.9) of logarithms of errors for $\rho = 1$

On the other hand, the top right $m \times n$ submatrices of HW and $\Theta\Omega$ should be the same, because of the zero elements of W. Therefore, instead of the second part of expression (7.4), we consider the logarithm

$$\log_{10} \max \{ \, | (\Theta\,\Omega)_{jk} | \, : \, 1 \le j \le m, \; m+1 \le k \le m+n \, \}. \tag{7.9}$$

Moreover, we retain the value (7.5) of σ_t. The values of the terms (7.8) and (7.9) for all the experiments of Table 1 are reported in Table 2, keeping the practice of placing the errors (7.8) of the Lagrange conditions before the solidus signs.

Many of the entries in Table 2 are less than the corresponding entries in Table 1, especially in the Test 3 and 4 rows and in the last column. Therefore another good reason for working with Θ instead of with H, as recommended in Section 6, is that the accuracy may be better. The automatic correction of the initial errors of the Lagrange conditions, shown in the Test 5 row of Table 2, is particularly welcome. This feature of the updating formula (6.11) was discovered by numerical experiments, which assisted the development of Lemma 3.

Reductions in ρ are made in Tests 6 and 7. Specifically, $\rho = 1$ is picked initially as before, and the changes to ρ are that it is decreased by a factor of 10 after every 500 iterations. Otherwise, all of the choices of the opening paragraph of this section are retained, the vector $\underline{x}_0$ being the zero vector and $10^{-4}\underline{e}$ in Tests 6 and 7, respectively. The purpose of these tests is to investigate the accuracy of the updating formulae when the interpolation points tend to cluster near the origin as ρ is decreased, so we require the way of selecting t on each iteration to provide the clustering automatically. Therefore the point $\underline{x}_t$ that is dismissed by expression (3.5) should be relatively far from the origin, provided that $|\ell_t(\underline{x}^+)|$ is not too small. These two conditions oppose each other when $\|\underline{x}_t\|/\rho$ is large, because then the positions of the interpolation points cause $|\ell_t(\underline{x})|$ to be of magnitude $(\rho/\|\underline{x}_t\|)^2$ in the neighbourhood $\{\underline{x} : \|\underline{x}\| \le \rho\}$, which is where $\underline{x}^+$ is generated. Therefore a technique that responds adequately to the quadratic decay of the Lagrange function has to allow $|\ell_t(\underline{x}^+)|$ to be much less than before. We counteract the quadratic decay by introducing a cubic term, letting t on each iteration be the integer i that maximizes the product

$$|\ell_i(\underline{x}^+)| \max\{\, 1,\, (\|\underline{x}_i\|/\rho)^3 \,\}, \qquad i = 1, 2, \ldots, m. \tag{7.10}$$

Thus it is usual in Tests 6 and 7 for all the conditions $\|\underline{x}_i\| \le \rho$, $i = 1, 2, \ldots, m$, to hold immediately before a reduction in ρ.

We continue to measure the accuracy of the Lagrange conditions by the first part of expression (7.4) or by the quantity (7.8) when working with H or Θ, respectively, It is no longer suitable, however, to consider the largest modulus of the errors in the constraints. In particular, the element $(HW)_{1\,m+1}$ is the sum $\sum_{k=1}^m H_{1k}$, so typical computer rounding errors may cause the modulus of this sum to be at least $10^{-16} \sum_{k=1}^m |H_{1k}|$. Further, because the elements (2.7) are $\mathcal{O}(\rho^4)$ when $\underline{x}_0$ is at the origin, the elements H_{jk}, $1 \le j, k \le m$, which are independent of $\underline{x}_0$ in theory, are of magnitude ρ^{-4}, as mentioned in the paragraph that includes the derivatives (5.6). Thus, when ρ reaches its final value of 10^{-7} in Tests 6 and 7, we expect the constraint errors to be at least 10^{12}, so consideration of the second part of expression (7.4) would be a misleading way of identifying good accuracy. Therefore we calculate the values

$$\left.\begin{array}{l} \log_{10} \max\{\, |\,(HW)_{jk}| \,/\, \sum_{i=1}^m |H_{ji} W_{ik}| \,:\, 1 \le j \le m,\ m+1 \le k \le m+n+1 \,\} \\ \log_{10} \max\{\, |\,(\Theta\,\Omega)_{jk}| \,/\, \sum_{i=1}^m |\Theta_{ji} \Omega_{ik}| \,:\, 1 \le j \le m,\ m+1 \le k \le m+n \,\} \end{array}\right\}, \tag{7.11}$$

instead of the terms on the right of the solidus signs in Tables 1 and 2.

The results of Tests 6 and 7 are given in Table 3, the numbers (7.11) being placed after the solidus signs. The presence of (H) and (Θ) in the first column distinguishes between the updating formulae of Sections 4 and 6, respectively. Each entry in the main part of the table is now the greatest value of the relevant logarithm of the error that occurs during a sequence of 1000 consecutive iterations. We make this change from Tables 1 and 2 in order to capture the largest errors, because they are sensitive to any particularly small values of $|\ell_t(\underline{x}^+)|$ that are admitted by our use of expression (7.10).

The Test 6 entries in Table 3 suggest that the accuracy of both updating methods is good when $\underline{x}_0$ is at the origin. Indeed, any errors in the conditions $\ell_j(\underline{x}_k) = \delta_{jk}$, $1 \le j, k \le m$, that arise from reductions in ρ are corrected by the stability properties of the updating formulae. A different feature prevents an unwelcome accumulation of errors in the quantities (7.11), namely that the denominators of these quantities grow as ρ is decreased. On the other hand, the results of Test 7 in the last two columns of Table 3 expose losses of accuracy that are unacceptable. Therefore we expect the work of Section 5 on changes to $\underline{x}_0$ to be required in practice. Some of the losses are due to severe cancellation in formulae (4.19) and (6.14) for β_t when $\|\underline{x}_0\|$ is much larger than ρ. Indeed, it is proved at the end of Section 5 that β_t is independent of $\underline{x}_0$, and the magnitude $\beta_t = \mathcal{O}(\rho^4)$ can be deduced

	Iterations 1–1000	Iterations 1001–2000	Iterations 2001–3000	Iterations 3001–4000
Test 6 (H)	$-12.9\,/\,-14.5$	$-12.6\,/\,-14.7$	$-12.7\,/\,-14.7$	$-12.7\,/\,-14.7$
Test 6 (Θ)	$-12.7\,/\,-14.3$	$-12.7\,/\,-14.6$	$-12.6\,/\,-14.4$	$-12.7\,/\,-14.6$
Test 7 (H)	$-12.7\,/\,-14.2$	$-12.4\,/\,-14.4$	$+7.9\,/\,-2.7$	$+10.0\,/\,-1.4$
Test 7 (Θ)	$-12.9\,/\,-14.2$	$-12.8\,/\,-14.5$	$-3.1\,/\,-9.5$	$+9.3\,/\,-4.9$

Table 3: Greatest logarithms of errors when ρ is decreased

similarly. It follows that, at the end of Test 7, where not only $\|\underline{x}_0\|=10^{-4}n^{1/2}$ but also $\rho=10^{-7}$ occur, the difference $\frac{1}{2}\|\underline{x}^+-\underline{x}_0\|^4-\underline{w}^T H\underline{w}$ in expression (4.19) should be approximately 10^{-28}, but the position of $\underline{x}_0$ with $n=50$ give the estimate $\frac{1}{2}\|\underline{x}^+-\underline{x}_0\|^4\approx 1.25\times 10^{-13}$. The cancellation is less, however, when β_t is obtained from formula (6.14), because the product $\|\underline{x}^{\rm mid}-\underline{x}_0\|^2\|\underline{d}\|^2$ is a dominant part of this formula, and the final value of the product is about $\|\underline{x}_0\|^2\rho^2=5\times 10^{-21}$. Other reasons for adjusting $\underline{x}_0$ occasionally are given in Section 5.

The kinds of difficulties that are overcome by the given updating procedures are shown by the magnitudes of the elements of H when ρ is very small. We estimate these magnitudes from the structure

$$W \;=\; \left(\begin{array}{c|c|c} \mathcal{O}(\rho^4) & \underline{e} & \mathcal{O}(\rho) \\ \hline \underline{e}^T & 0 & 0 \\ \hline \mathcal{O}(\rho) & 0 & 0 \end{array}\right) \tag{7.12}$$

of the matrix (5.1), when $\underline{x}_0$ and the interpolation points satisfy $\|\underline{x}_i-\underline{x}_b\|\le\rho$, $i=0,1,\ldots,n$. Because of the theoretical identity $HW=I$, all of the nonzero products in the set $\{H_{ji}W_{ik} : i=1,2,\ldots,m+n+1\}$ may have similar magnitudes, where j and k are any integers from $[1,m+n+1]$, and the magnitudes should be about one in the cases with $j=k$. Therefore we expect the matrix $H=W^{-1}$, which is symmetric, to have the form

$$H \;=\; \left(\begin{array}{c|c|c} \mathcal{O}(\rho^{-4}) & \mathcal{O}(1) & \mathcal{O}(\rho^{-1}) \\ \hline \mathcal{O}(1) & \mathcal{O}(\rho^4) & \mathcal{O}(\rho^3) \\ \hline \mathcal{O}(\rho^{-1}) & \mathcal{O}(\rho^3) & \mathcal{O}(\rho^2) \end{array}\right). \tag{7.13}$$

The $\mathcal{O}(\rho^{-4})$ elements have been mentioned already, and all the magnitudes (7.13) are supported by the numerical results at the end of Test 6. Indeed, the value of ρ is 10^{-7}, and the least and greatest moduli of elements in each partition of expression (7.13) are as follows. The $\mathcal{O}(\rho^{-4})$ and $\mathcal{O}(\rho^{-1})$ terms are in the intervals $[1.9{\times}10^{23}, 1.6{\times}10^{28}]$ and $[6.0{\times}10^{2}, 4.6{\times}10^{6}]$, respectively, the $\mathcal{O}(1)$, $\mathcal{O}(\rho^2)$ and $\mathcal{O}(\rho^3)$ terms are in $[1.6{\times}10^{-6}, 3.3{\times}10^{-2}]$, $[2.1{\times}10^{-19}, 6.5{\times}10^{-15}]$ and $[5.6{\times}10^{-26}, 1.7{\times}10^{-23}]$, while the single $\mathcal{O}(\rho^4)$ element of H has the modulus 1.7×10^{-30}. On the other hand, at the beginning of Test 6 when ρ is one, the nonzero elements of H are of magnitude one, and all the elements of H acquire this property during the first 500 iterations of the test, after the interpolation points have been moved from their initial positions, which cause many elements of H to be zero at the beginning of the calculation.

Therefore major changes are made to H during Test 6 (H) by the sequence of applications of the updating formula (4.12). On the other hand, updating methods that require the matrices to be well-conditioned would not be suitable when ρ has the value 10^{-7} in expression (7.13). If we take the view that we are calculating each H because we wish to solve the system of equations (2.8), then we are updating the inverse of the matrix of the system, but it is not unusual to read that unnecessary errors may occur if one works with inverses, so the updating of matrix factorizations is often recommended instead. Our theoretical and numerical results, however, provide strong support for the use of inverse matrices. Further, the elements of these matrices give the coefficients of the Lagrange functions that are introduced in Section 3. We welcome the accuracy that is achieved by our methods. In particular, because only the first m rows of $HW-I$ contribute to Table 3, we note now that the final matrix H of the Test 6 (H) calculation has the property

$$|\,(HW)_{jk} - \delta_{jk}\,| \Big/ \sum_{i=1}^{m+n+1} |H_{ji}W_{ik}| \;\leq\; 4.8\times 10^{-14}, \quad 1\leq j,k\leq m+n+1. \qquad (7.14)$$

Moreover, each application of formula (4.12) or (6.11) requires only $\mathcal{O}(\{m+n\}^2)$ computer operations. Therefore we expect these formulae to become very useful in practice. The author has begun to develop Fortran software for general unconstrained minimization calculations without derivatives that employs the given techniques.

Acknowledgement

The author is very grateful to the referees. They provided many suggestions that improved the presentation of this work.

References

C.G. Broyden, J.E. Dennis and J.J. Moré (1973), "On the local and superlinear convergence of quasi-Newton methods", *J. Inst. Math. Appl.*, Vol. 12, pp. 223–245.

A.R. Conn, N.I.M. Gould and Ph.L. Toint (2000), *Trust-Region Methods*, MPS–SIAM Series on Optimization (Philadelphia).

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

R. Fletcher (1987), *Practical Methods of Optimization*, John Wiley & Sons (Chichester).

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

M.J.D. Powell (2001), "On the Lagrange functions of quadratic models that are defined by interpolation", *Optim. Meth. Software*, Vol. 16, pp. 289–309.

M.J.D. Powell (2002), "UOBYQA: unconstrained optimization by quadratic approximation", *Math. Programming*, Vol. 92, pp. 555–582.

M.J.D. Powell (2003), "On trust region methods for unconstrained minimization without derivatives", Report No. DAMTP 2002/NA02, University of Cambridge (accepted for publication in *Math. Programming*).