Research paper

Struct Multidisc Optim 27, 326–336 (2004)
DOI 10.1007/s00158-004-0391-2

Multipoint cubic surrogate function for sequential approximate optimization*

R.A. Canfield

Abstract Multipoint cubic approximations are investigated as surrogate functions for nonlinear objective and constraint functions in the context of sequential approximate optimization. The proposed surrogate functions match actual function and gradient values, including the current expansion point, thus satisfying the zero and first-order necessary conditions for global convergence to a local minimum of the original problem. Function and gradient information accumulated from multiple design points during the iteration history is used in estimating a reduced Hessian matrix and selected cubic terms in a design subspace appropriate for problems with many design variables. The resulting approximate response surface promises to accelerate convergence to an optimal design within the framework of a trust region algorithm. The hope is to realize computational savings in solving large numerical optimization problems. Numerical examples demonstrate the effectiveness of the new multipoint surrogate function in reducing errors over large changes in design variables.

Key words multipoint approximation, optimization, response surface, surrogate function

Received: 25 March 2003
Revised manuscript received: 6 November 2003
Published online: 29 June 2004

R.A. Canfield

Air Force Institute of Technology, Dept. of Aeronautical and Astronautical Engineering, AFIT/ENY, 2950 Hobson Way, WPAFB, Ohio 45433–7765, USA
e-mail: Robert.Canfield@afit.edu

* Presented as Paper 2002-5575 at the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Atlanta, Georgia, September 4–6, 2002. Revised version of the paper presented at the Third ISSMO/AIAA Internet Conference on Approximations in Optimization October 14–25, 2002.

1 Introduction

A nonlinear optimization problem may be defined as finding the design variable vector $\mathbf{x}^*$ that minimizes an objective function $f : \Re^n \to \Re$ over the feasible domain $\Omega \in \Re^n$

$$f(\mathbf{x}^*) = \min_{\mathbf{x}\in\Omega} f(\mathbf{x})$$
$$\Omega = \{\mathbf{x} \mid \mathbf{g}(\mathbf{x}) \le 0,\ \mathbf{h}(\mathbf{x}) = 0\} \tag{1}$$

where the feasible region is the set of all design variables that satisfy vector inequality constraints $\mathbf{g}(\mathbf{x})$ and vector equality constraints $\mathbf{h}(\mathbf{x})$. The vector inequality is taken to mean that each and every component of the vector must satisfy the inequality. Side constraints on the design variables may be incorporated in the inequality constraints $\mathbf{g}(\mathbf{x})$.

Although many numerical methods are readily available to solve this problem, we are concerned with developing as efficient an algorithm as possible for the case where function and gradient evaluations of the objective and constraints may be expensive. Such is the case in many practical design problems involving computation structural mechanics, computational fluid dynamics (CFD), and robust control, where a complete finite element or CFD analysis or dynamic simulation may be required, for example, in order to evaluate the functions. To this end, various approximation concepts have been proposed (Schmit and Miura 1977). Initial efforts focussed on selecting appropriate intermediate design variables (Haftka and Gurdal 1992). More recently, researchers have examined intermediate response quantities that may be approximated more accurately than the actual objective or constraint functions (Haftka and Gurdal 1992; Barthelemy and Haftka 1993).

To illustrate these approximation concepts, let $\mathbf{y}(\mathbf{x})$ be a vector of intermediate design variables that can be expressed explicitly in terms of the original design variables $\mathbf{x}$, e.g.reciprocal variables $y_i = 1/x_i$. Let $\mathbf{q}(\mathbf{y})$ be a vector of intermediate response quantities from which the objective and constraint functions can be calculated explicitly

$$f(\mathbf{x}) = \hat{f}(\mathbf{q}(\mathbf{y}(\mathbf{x})), \mathbf{y}(\mathbf{x}), \mathbf{x}), \tag{2a}$$

$$\mathbf{g}(\mathbf{x}) = \hat{\mathbf{g}}(\mathbf{q}(\mathbf{y}(\mathbf{x})), \mathbf{y}(\mathbf{x}), \mathbf{x}), \tag{2b}$$

$$\mathbf{h}(\mathbf{x}) = \hat{\mathbf{h}}(\mathbf{q}(\mathbf{y}(\mathbf{x})), \mathbf{y}(\mathbf{x}), \mathbf{x}). \tag{2c}$$

In the approximation concepts approach, an exact analysis is carried out to calculate the response quantities $\mathbf{q}(\mathbf{y})$. Because this is typically quite expensive, the response quantities are approximated, usually by a linear Taylor series in the intermediate variables $\mathbf{y}$.

$$\widetilde{\mathbf{q}}_L(\mathbf{y}) = \mathbf{q}(\mathbf{y}_o) + \{\nabla \mathbf{q}(\mathbf{y}_o)\}^T \{\mathbf{y} - \mathbf{y}_o\} \tag{3}$$

Approximations to the original functions can be made easily, as they are explicit functions of the intermediate responses formed by substituting (3) into (2) in place of the actual response quantities

$$\widetilde{f}(\mathbf{x}) = \hat{f}(\widetilde{\mathbf{q}}_L(\mathbf{y}(\mathbf{x})), \mathbf{y}(\mathbf{x}), \mathbf{x}), \tag{4a}$$

$$\widetilde{\mathbf{g}}(\mathbf{x}) = \hat{\mathbf{g}}(\widetilde{\mathbf{q}}_L(\mathbf{y}(\mathbf{x})), \mathbf{y}(\mathbf{x}), \mathbf{x}), \tag{4b}$$

$$\widetilde{\mathbf{h}}(\mathbf{x}) = \hat{\mathbf{h}}(\widetilde{\mathbf{q}}_L(\mathbf{y}(\mathbf{x})), \mathbf{y}(\mathbf{x}), \mathbf{x}). \tag{4c}$$

Now the approximate sub-problem is defined as

$$\widetilde{f}(\mathbf{x}^*) = \min_{\mathbf{x} \in \widetilde{\Omega}} \widetilde{f}(\mathbf{x})$$

$$\widetilde{\Omega} = \left\{ \mathbf{x} \mid \widetilde{\mathbf{g}}(\mathbf{x}) \leq 0,\ \widetilde{\mathbf{h}}(\mathbf{x}) = 0 \right\} \tag{5}$$

where $\widetilde{\Omega}$ is the domain of the approximate feasible region. Equations (5) are solved iteratively until the approximations in (4) converge to the exact values given by (2) to within an acceptable tolerance.

The iterative solution of this approximate sub-problem can drastically reduce the number of full exact analyses, provided the approximations are of high enough quality. Nevertheless, unlike certain unconstrained optimization methods such as conjugate gradient or variable metric methods, which accumulate historical information about a function from previous design points, the aforementioned approximation concepts are based on a single design point. New approximations are reconstructed each time an exact analysis is made. Information from previous analyses is typically lost. Some methods do incorporate information from previous iterations into the approximations. Haftka and Gurdal (1992) classify such multipoint approximations as *mid-range approximations*. Haftka *et al.* (1987) showed that two- and three-point approximations can be accurate for interpolation, but not necessarily for extrapolation. Fadel *et al.* (1990) incorporated previous gradient information in a two-point approximation that was used to select better intermediate design variables for a linear approximation. In contrast, Rasmussen (1990) incorporated only previous function values, while ignoring gradients from previous iterations. He demonstrated stable convergence for a cantilever box beam example for which other methods failed to converge; however, he observed that more iterations were required than for the method of moving asymptotes (MMA), a single-point approximation (Svanberg 1987). Snyman and Stander (1994) also used only the previous function value in creating a two-point quadratic approximation. Toropov *et al.* (1993) proposed a simulation approach to structural optimization in which multiple points generated as a "plan of experiments" are used as the basis for multipoint approximations. He presented results comparable in efficiency to Svanberg's single-point MMA. Xu and Grandhi (1998) developed a two-point adaptive nonlinear approximation (TANA2) that incorporates Fadel's intermediate variables for gradient matching with Snyman and Stander's quadratic correction for function matching at a single previous point. Xu and Grandhi (2000) extended TANA2 to multiple points through the blending functions.

Vanderplaats (1984) outlines an approach estimating certain terms in the Hessian matrix based on previous information; however, details are not provided. For example, it appears that only a subset of Hessian terms would be approximated, but no guidance is available for choosing that subset. In a different vein, Miura and Schmit (1978) demonstrated the higher accuracy of quadratic Taylor series approximations for structural optimization with frequency constraints. However, the computational savings due to improved convergence were balanced by the increased cost of calculating second derivatives. The present approach fully utilizes information from previous iterations, in the same spirit as quasi-Newton methods, which gradually build an approximate Hessian (Gill *et al.* 1981).

Canfield (1994) derived a multipoint quadratic approximation (MQA). The objective was to extend the accuracy of a class of Taylor series approximations without having to calculate the second-order derivatives appearing in the Hessian matrix. Instead, quadratic correction terms were calculated using function and gradient values from previous iterates. The resulting MQA succeeded in exactly matching previously known function values at nearby points. It matched function values at more distant points and previously known gradient values as closely as possible. The obvious drawback for large design problems was that the approximate Hessian matrix it generated grew as the square of the number of design variables. Thus, not only the calculation of the MQA itself, but also its use in the solution of the approximate optimization sub-problem, becomes too expensive as the problem size grows. Canfield (2001) addressed these deficiencies by deriving a reduced Hessian approximation, created in a design subspace spanning all previous points.

The proposed *multipoint cubic approximation* (MCA) builds upon the MQA by eliminating the pseudo-inverse approach to matching previous gradient values. Instead, cubic terms in the reduced design space are used. In this way enough coefficients are determined so that all previous information, both function and gradient values, are

reproduced. The goal of this research, then, is to garner the higher accuracy and concomitant improvement in convergence of a Taylor series approximation without incurring the expense of calculating second or third derivatives. The current approach satisfies the first-order consistency conditions required for global convergence of approximate models (Alexandrov *et al.* 1998). The sequential solution of successive MCA models relies on a model trust region approach (Fletcher 1981) applied to an augmented Lagrangian function (Conn *et al.* 1991). The present approach employed a trust region strategy similar to (Wujek and Renaud 1998), although a one-dimensional line search was an option to determine the adaptive move limits, as suggested by Fletcher (1981).

2 Theory

In this section the notation for a multivariate function $f(\mathbf{x})$ will be used for convenience. Nevertheless, the approximation is not limited to the objective function. Indeed, in the context of the approximate sub-problem, the functions being approximated would normally be elements of the vector $\mathbf{q}(\mathbf{y})$. However, to illustrate the generality of the MCA, intermediate responses were *not* used in the numerical examples of Sect. 3.

2.1 Using previous information

Consider a quadratic approximation $\widetilde{f}_Q$ to the function $f(\mathbf{x})$

$$\widetilde{f}_Q(\mathbf{x},\mathbf{H}) = f(\mathbf{x}_k) + \{\nabla\mathbf{f}(\mathbf{x}_k)\}^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T\mathbf{H}\,\Delta\mathbf{x} \qquad (6)$$

where $\mathbf{x}_k$ is the design vector at the kth iteration, $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_k$ is the change in the design vector, $\nabla\mathbf{f}(\mathbf{x}_k)$ is the function's gradient, and $\mathbf{H}$ is an approximation to its Hessian matrix. If $\mathbf{H}$ were composed of analytic second derivatives, (6) would represent a second-order Taylor series approximation. However, instead of evaluating the second derivatives at the current design point $\mathbf{x}_k$, the Hessian matrix will be estimated from function and gradient values at previous points. In the first iteration, of course, there is no previous information, so a standard linear Taylor series or first-order conservative approximation is employed.

Let us consider the elements of the Hessian matrix $\mathbf{H}$ as unknown parameters. We would like the quadratic approximation (6) to reproduce known function values from the set of $k-1$ previous design points

$$\widetilde{f}_Q(\mathbf{x}_p,\mathbf{H}) = f(\mathbf{x}_p)\,, \quad p = 1,2,\ldots,k-1 \qquad (7)$$

where the symmetric Hessian matrix has $N = n(n+1)/2$ unknown elements for n design variables. We would also like to reproduce the gradients from previous design points as closely as possible. The gradients of the quadratic approximation (6) evaluated at the previous design points should satisfy

$$\nabla_{\mathbf{x}}\,\widetilde{\mathbf{f}}_Q(\mathbf{x}_p,\mathbf{H}) = \nabla\mathbf{f}(\mathbf{x}_k) + \mathbf{H}(\mathbf{x}_p - \mathbf{x}_k) =$$

$$\nabla\mathbf{f}(\mathbf{x}_p)\,, \quad p = 1,2,\ldots,k-1\,. \qquad (8)$$

where (7) and (8) constitute a set of linear equations in the N unknown elements of the Hessian matrix $\mathbf{H}$; however, the equations are in general inconsistent. A Hessian satisfying (8) will not satisfy (7) in general, when $f(\mathbf{x})$ is not a quadratic function. Moreover, for a large number of design variables, it is impractical to solve for so many Hessian terms. These issues will be addressed in the next section.

2.2 Reduced basis multipoint cubic approximation

The preceding derivation is now modified so that the approximate Hessian is limited to a subspace of the design variables determined by the vectors connecting the current design point to the previous design points where function and gradient information is available. Thus, the reduced-order Hessian approximation has a dimension equal to the number of iterates traversed in the optimization solution (or is predetermined by a design of experiments sampling). In addition, cubic terms in the reduced space are appended in order to provide enough freedom to match both previous functions and gradients.

We seek a reduced-order Hessian approximation. Let the difference between the current base point, $\mathbf{x}_k$, and previous points, $\mathbf{x}_p$, be arranged as columns in a matrix

$$\mathbf{X}_k = [\{\mathbf{x}_1 - \mathbf{x}_k\}\ \{\mathbf{x}_2 - \mathbf{x}_k\}\ \ldots\ \{\mathbf{x}_{k-1} - \mathbf{x}_k\}]\,. \qquad (9)$$

Next, we transform the matrix of changes from previous iterations to an orthonormal basis, using the Grahm-Schmidt orthonormalization procedure, $\mathbf{G}_k = G(\mathbf{X}_k)$, so that the previous points may be represented as

$$\mathbf{x}_p = \mathbf{x}_k + \mathbf{G}_k\mathbf{d}_p \qquad (10)$$

where $\mathbf{d}_p = \mathbf{d}(\mathbf{x}_p)$ is the vector of reduced basis coefficients corresponding to a previous point, $\mathbf{x}_p$. Now the basis vectors may be used to project any design vector, $\mathbf{x}$, onto the subspace of the current and previous points.

$$\mathbf{d}(\mathbf{x}) = \mathbf{G}_k^T\{\mathbf{x} - \mathbf{x}_k\}\,. \qquad (11)$$

The *multipoint cubic approximation* (MCA) is formed by adding quadratic and cubic terms in the subspace to the linear approximation in the full space,

$$\widetilde{f}_C(\mathbf{x}) = \widetilde{f}_L(\mathbf{x}) + \frac{1}{2}\mathbf{d}(\mathbf{x})^T\widetilde{\mathbf{H}}\,\mathbf{d}(\mathbf{x}) + \frac{1}{6}\sum_{i=1}^{P}\beta_i d_i^3(\mathbf{x}) \qquad (12)$$

where d_i is the ith element of the vector $\mathbf{d}$ defined in (11) and β_i represent the unknown cubic coefficients to be determined together with the Hessian approximation $\widetilde{\mathbf{H}}$ in the proposed MCA. The linear approximation is denoted as

$$\widetilde{f}_L(\mathbf{x}) = f(\mathbf{x}_k) + \{\nabla \mathbf{f}(\mathbf{x}_k)\}^T \{\mathbf{x} - \mathbf{x}_k\} . \tag{13}$$

The size of the reduced Hessian approximation $\widetilde{\mathbf{H}}$ is just the square of the number of previous points. Likewise, there are only as many cubic terms as the number of previous points, $k-1$.

The determination of the reduced-order approximation occurs in the subspace spanned by $\mathbf{X}_k$ instead of $\mathbf{x}$. The conditions to be satisfied in determining $\widetilde{\mathbf{H}}$ and $\boldsymbol{\beta}$ are the function-matching conditions

$$\widetilde{f}_C(\mathbf{x}_p) = f(\mathbf{x}_p) , \quad p = 1, 2, \ldots , k-1 \tag{14}$$

and the gradient-matching conditions

$$\nabla \widetilde{\mathbf{f}}_C(\mathbf{x}_p) = \nabla \mathbf{f}(\mathbf{x}_p), \quad p = 1, 2, \ldots , k-1 . \tag{15}$$

The left side of (14) is found by evaluating (12) at each previous point, $\mathbf{x}_p$. To evaluate the left side of (15) requires the differentiation of (12) with respect to the design vector, $\mathbf{x}$.

$$\nabla \widetilde{\mathbf{f}}_C(\mathbf{x}) = \nabla \mathbf{f}(\mathbf{x}_k) + \frac{1}{2}\mathbf{G}_k \left(\widetilde{\mathbf{H}} + \widetilde{\mathbf{H}}^T\right) \mathbf{G}_k^T (\mathbf{x} - \mathbf{x}_k) +$$

$$\frac{1}{2} \sum_{i=1}^{P} \beta_i d_i^2(\mathbf{x}) \frac{\partial d_i}{\partial \mathbf{x}} \tag{16}$$

After substituting (11) into (16), pre-multiplying by the transpose of the orthonormal transformation, and recognizing that each $\frac{\partial d_i}{\partial \mathbf{x}}$ is the ith column of the transformation matrix $\mathbf{G}_k$, we have

$$\mathbf{G}_k^T \left\{\nabla \widetilde{\mathbf{f}}_C(\mathbf{x}) - \nabla \mathbf{f}(\mathbf{x}_k)\right\} = \frac{1}{2} \left(\widetilde{\mathbf{H}} + \widetilde{\mathbf{H}}^T\right) \mathbf{d} +$$

$$\frac{1}{2} \left[\text{diag}\left(\mathbf{d}^2\right)\right] \boldsymbol{\beta} \tag{17}$$

where diag $(\mathbf{d}^2)$ indicates a diagonal matrix of the squares of the reduced basis coefficients, d_i^2. The left side of (17) is the difference between the MCA gradient and the actual gradient at the current point, projected onto the subspace. Equating the MCA gradient to the actual gradient at each previous point according to (15), substituting into (17), and assuming symmetry of $\widetilde{\mathbf{H}}$ gives

$$\boldsymbol{\gamma}_p = \widetilde{\mathbf{H}} \mathbf{d}_p + \frac{1}{2} \left[\text{diag}\left(\mathbf{d}_p^2\right)\right] \boldsymbol{\beta} \tag{18}$$

where

$$\boldsymbol{\gamma}_p \equiv \mathbf{G}_k^T \{\nabla \mathbf{f}(\mathbf{x}_p) - \nabla \mathbf{f}(\mathbf{x}_k)\} \tag{19}$$

is the projected gradient difference, i.e.the difference between the gradient at the previous and current points projected onto the subspace. Pre-multiplying (18) by $\frac{1}{2}\mathbf{d}_p^T$ and rearranging terms allows us to isolate the quadratic terms in order to eliminate them from the function-matching equations that follow.

$$\frac{1}{2}\mathbf{d}_p^T \widetilde{\mathbf{H}} \mathbf{d}_p = \frac{1}{2}\mathbf{d}_p^T \boldsymbol{\gamma}_p - \frac{1}{4}\mathbf{d}_p^T \left[\text{diag}\left(\mathbf{d}_p^2\right)\right] \boldsymbol{\beta} \tag{20}$$

Introducing (20) into (12), which is equated to the previous function values according to (14), provides a system of equations to solve for the vector $\boldsymbol{\beta}$ of cubic coefficients. Denoting δ_p as the difference between the linear approximation (13) and the actual function value at point p, (14) becomes

$$\delta_p \equiv f(\mathbf{x}_p) - \widetilde{f}_L(\mathbf{x}_p) = \frac{1}{2}\mathbf{d}_p^T \widetilde{\mathbf{H}} \mathbf{d}_p + \frac{1}{6} \left\{\mathbf{d}_p^3\right\}^T \boldsymbol{\beta} \tag{21}$$

where $\mathbf{d}_p^3$ is a column vector of the reduced basis coefficients cubed, $d_{p_i}^3$, for a previous point, $\mathbf{x}_p$. Next, substitute (20) into (21) to eliminate the Hessian matrix, and solve for the cubic coefficients

$$\boldsymbol{\beta} = -12 \left[\mathbf{D}^3\right]^{-T} \bar{\boldsymbol{\delta}} \tag{22}$$

where $\mathbf{D}^3$ represents a matrix with columns $\mathbf{d}_p^3$ and each element of the column vector $\bar{\boldsymbol{\delta}}$ is defined as

$$\bar{\delta}_p = \delta_p - \frac{1}{2}\boldsymbol{\gamma}_p^T \mathbf{d}_p . \tag{23}$$

Once $\boldsymbol{\beta}$ is solved from (22), it is substituted into (18), which is a vector equation for each previous point. Rearranging the result yields

$$\widetilde{\mathbf{H}} \mathbf{d}_p = \boldsymbol{\gamma}_p - \frac{1}{2} \left[\text{diag}\left(\mathbf{d}_p^2\right)\right] \boldsymbol{\beta} \equiv \bar{\boldsymbol{\gamma}}_p \tag{24}$$

where $\bar{\boldsymbol{\gamma}}_p$ defines a gradient difference error. Arranging the vectors in (24) column-wise, for each previous point, in a single matrix equation permits the following solution for the reduced approximate Hessian matrix

$$\widetilde{\mathbf{H}} = \boldsymbol{\Gamma}[\mathbf{D}]^{-1} \tag{25}$$

where $\mathbf{D}$ represents a matrix with columns $\mathbf{d}_p$ and

$$\boldsymbol{\Gamma} = \left[\{\bar{\boldsymbol{\gamma}}_1\} \{\bar{\boldsymbol{\gamma}}_1\} \ldots \{\bar{\boldsymbol{\gamma}}_{k-1}\}\right] \tag{26}$$

is a matrix composed of column vectors of the gradient difference errors. The matrix $\mathbf{D}$ is guaranteed to be full rank by virtue of the orthonormalization. It can be made square by retaining only as many previous function values as there are independent basis vectors in $\mathbf{X}_k$, i.e. the rank (number of columns) of $\mathbf{G}_k$. Also, although $\widetilde{\mathbf{H}}$ was assumed symmetric in going from (17) to (18), the solution of (25) does not enforce symmetry. Therefore, there will be some error in matching previous gradient information using (15).

In summary, substituting the MCA $\widetilde{f}_C(\mathbf{x})$ of (12), constructed in the reduced orthonormal basis $\mathbf{G}_k$ associated

with (9), into the function- and gradient-matching conditions of (14) and (15) leads to (22) and (25) for the cubic and quadratic coefficients, $\boldsymbol{\beta}$ and $\widetilde{\mathbf{H}}$, respectively. Solution of the latter involves the projected gradient differences $\boldsymbol{\gamma}_p$, given by (19), and the linear approximation error for function values δ_p, given by (21).

2.3 Enriched reduced basis

The careful reader will observe that the transformation from (15) to (18) implies that only the projected gradient differences $\boldsymbol{\gamma}_p$ of (19) are matched, not the original gradients $\nabla\widetilde{\mathbf{f}}_C(\mathbf{x}_p)$ specified in (15). Whether the gradient differences can be matched in (16) depends on whether the gradient differences are spanned by the reduced basis $\mathbf{X}_k$. To recover the original gradients, the gradient differences $\{\nabla\mathbf{f}(\mathbf{x}_p) - \nabla\mathbf{f}(\mathbf{x}_k)\}$ can be appended to the design vector differences in the $\mathbf{X}_k$ basis. The drawback to this *enriched reduced basis* is that each function has a different reduced basis. This modification did not produce a significant difference in the optimization results for the examples presented in Sect. 3. Before describing those examples, the procedure outlined by (1–5) for sequential approximate optimization is described next for the case when MCAs are used for the surrogate model (4).

2.4 Sequential MCA

The original optimization problem defined by (1) was replaced by a sequence of approximate sub-problems given by (5). In the first iteration the conservative approximation (single-point linear approximation with mixed direct and reciprocal variables, see (Barthelemy and Haftka 1993)) was used for the objective and constraint functions. At each subsequent iteration $k > 1$ the nonlinear objective function was approximated by a multipoint cubic approximation

$$\widetilde{f}_C^{(k)}(\mathbf{x}) = \widetilde{f}_L^{(k)}(\mathbf{x}) + \frac{1}{2}\mathbf{d}^T(\mathbf{x})\widetilde{\mathbf{H}}_f^{(k)}\,\mathbf{d}(\mathbf{x}) + \frac{1}{6}\sum_{i=1}^{k-1}\beta_{f_i}^{(k)} d_i^3(\mathbf{x}) \tag{27}$$

where the vector $\boldsymbol{\beta}_f^{(k)}$ of length $k-1$ was solved from (22) and the $(k-1)\times(k-1)$ matrix $\widetilde{\mathbf{H}}_f^{(k)}$ was solved from (25). Likewise, the nonlinear constraints were approximated in the same way. No intermediate responses were used for $\mathbf{g}(\mathbf{x})$. The reduced basis in (11) is not needed after the number of linearly independent previous points exceeds the number of design variables, n. Then, only the closest n previous design points $\mathbf{x}_p$ need be retained. Now $\mathbf{G}_k$ is an identity matrix and $\mathbf{d}$ simply becomes the set of closest linearly independent $\mathbf{x}$. Alternatively, more previous points can be retained, if the inverses in (22) and (25) are made pseudo-inverses.

The model trust region was applied to the augmented Lagrangian as a merit function

$$\Phi = f(\mathbf{x}) + \boldsymbol{\lambda}^T\boldsymbol{\psi} + R_p\boldsymbol{\psi}^T\boldsymbol{\psi} \tag{28}$$

for inequality constraints (Vanderplaats 1984) where

$$\psi_j = \max\left(g_j(\mathbf{x}), \frac{-\lambda_j}{2R_p}\right). \tag{29}$$

The *trust region ratio* is defined as the ratio of the actual reduction in the merit function to the merit reduction predicted by the MCA model (Fletcher 1981)

$$r^{(k)} = \frac{\Phi_k - \Phi_{k-1}}{\widetilde{\Phi}_k - \widetilde{\Phi}_{k-1}} \tag{30}$$

where the approximate merit function is

$$\widetilde{\Phi}_k = \widetilde{f}_C^{(k)}(\mathbf{x}) + \boldsymbol{\lambda}^T\widetilde{\boldsymbol{\psi}} + R_p\widetilde{\boldsymbol{\psi}}^T\widetilde{\boldsymbol{\psi}} \tag{31}$$

and the MCA is used to model not only the objective, $\widetilde{f}_C$, but also the active constraints, $\widetilde{\mathbf{g}}_C(\mathbf{x})$, through

$$\widetilde{\psi}_j = \max\left(\widetilde{g}_{Cj}^{(k)}(\mathbf{x}), \frac{-\lambda_j}{2R_p}\right). \tag{32}$$

Side constraints were adjusted each iteration by an adaptive move limit strategy governed by the MCA model trust region ratio (Fletcher 1981; Wujek and Renaud 1998). The trust region was reduced whenever $r^{(k)} < 0.25$ and it was expanded when $r^{(k)} > 0.75$.

A positive semi-definite approximation is desirable for inequality constraints, even though the constraint functions may not be positive semi-definite. The effect of a sign-indefinite Hessian approximation and cubic terms with mixed signs is to afford the optimizer an opportunity to compensate for constraint violations by seeking search directions that create large negative quadratic and cubic terms to achieve feasibility. The following modification was made for constraint approximations to compensate for this tendency.

$$\widetilde{g}_{Cj}^{(k)}(\mathbf{x}) = \widetilde{g}_{Lj}^{(k)}(\mathbf{x}) + \max\left(0,\ \frac{1}{2}\mathbf{d}^T(\mathbf{x})\widetilde{\mathbf{H}}_j^{(k)}\,\mathbf{d}(\mathbf{x}) + \frac{1}{6}\sum_{i=1}^{k-1}\beta_{ji}^{(k)} d_i^3(\mathbf{x})\right),$$
$$j = 1, \ldots, n_a \tag{33}$$

where n_a is the number of active constraints as determined by the current estimate of Lagrange multipliers, $\boldsymbol{\lambda}$. Inactive constraints were approximated by single-point linear functions (direct, reciprocal or conservative).

Application of move limits to an initially infeasible design may create an approximate sub-problem with no solution. To ensure that each sub-problem had a solution, slack variables were subtracted from all initially infeasible constraints in (2b). The non-negative slack variables were forced to zero by adjoining them to the approximate objective function given in (4a).

3
Numerical optimization

Results using the sequential MCA with a trust region strategy are presented. Iteration histories of the objective and constraint function values are shown for the present method and the sequential quadratic programming (SQP) method in MatLab (Branch and Andrew 1996).

3.1
Five-element cantilever beam optimization

The first example is a thin-walled, box beam, modelled with finite elements, loaded at the tip (Fig. 1 inset). The cantilever beam subject to a tip load is modelled with five Euler–Bernoulli finite elements (Svanberg 1987). The structural sizing optimization problem is to minimize the beam's mass subject to a maximum allowable tip deflection. The design variables are the width (height) of each element's square section for a box beam with fixed thickness. The objective function, volume of material in the beam, was minimized subject to a constraint on the tip deflection. The objective and constraint functions can be expressed as explicit functions of the design variables. Then the problem is to minimize

$$f(\mathbf{x}) = 0.0624(x_1 + x_2 + x_3 + x_4 + x_5)$$

subject to the constraint

$$g(\mathbf{x}) = 61/x_1^3 + 37/x_2^3 + 19/x_3^3 + 7/x_4^3 + 1/x_5^3 - 1 \leq 0\,.$$

The initial design is $x_i = 5$ for $i = 1, 2, 3, 4, 5$. The (exact) linear approximation was always used for the objective function, while linear, reciprocal and multipoint cubic approximations were tested for the constraint. The termination criteria were that the objective and constraint be

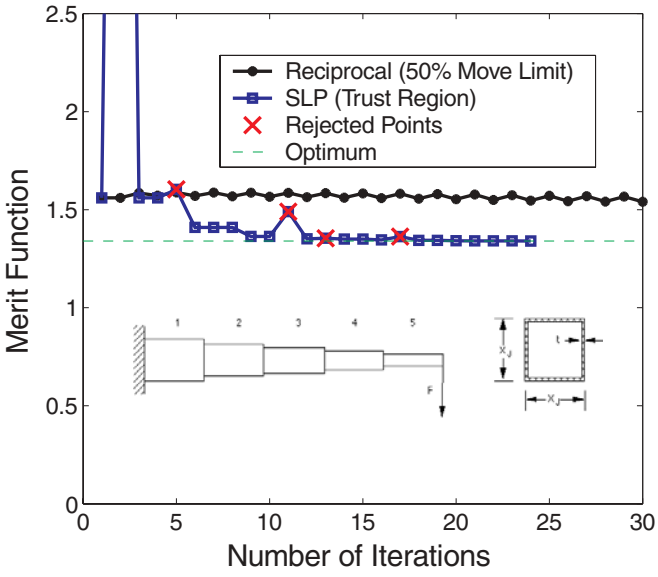

Fig. 1 Svanberg beam

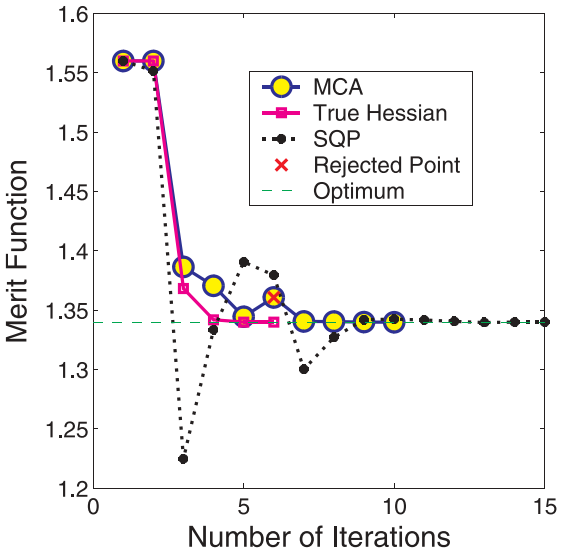

Fig. 2 Svanberg beam – multipoint cubic approximation

Table 1 MCA convergence for Svanberg beam

Iter. k	Object. f	Constraint g	Merit Φ	L.M. λ	Trust ratio $r^{(k)}$
1	1.560	0	1.5600	0.30	0.000
2	1.266	0.3924	1.5600	0.17	0.245
3	1.266	0.1977	1.3865	0.32	0.591
4	1.325	0.09748	1.3705	0.33	0.258
5	1.330	0.03145	1.3446	0.42	0.680
6	1.332	0.0793	1.3608	0.25	−1.281
7	1.335	0.01252	1.3406	0.43	0.504
8	1.338	0.004237	1.3404	0.45	0.092
9	1.340	0.0001055	1.3400	0.45	0.901
10	1.340	1.464×10^{-5}	1.3400	0.45	0.929

within 0.01% of the analytical solution. All model trust regions were initiated with 50% move limits. Fletcher's simple trust region adjustment strategy (Fletcher 1981) was used (no one-dimensional search to adapt the move limit).

Svanberg (1987) demonstrated that conventional first-order approximations with fixed move limits oscillate and do not converge to the optimum, as seen in Fig. 1 for fixed 50% move limits. Furthermore, although sequential linear programming (SLP) can be made globally convergent with a trust region strategy (it satisfies first-order consistency conditions), it converges slowly (also shown in Fig. 1).

In Fig. 2 the MCA is compared to MatLab's SQP and to the trust region algorithm applied to a second-order Taylor series approximation using the true Hessian of the constraint function. In principal, the latter is the most efficient quadratic solution possible with the present augmented Lagrangian approach. The merit function plotted for SQP is simply the objective function, not an augmented Lagrangian. The sequential MCA converged in 10 iterations, compared to 15 iterations for SQP and 6 it-

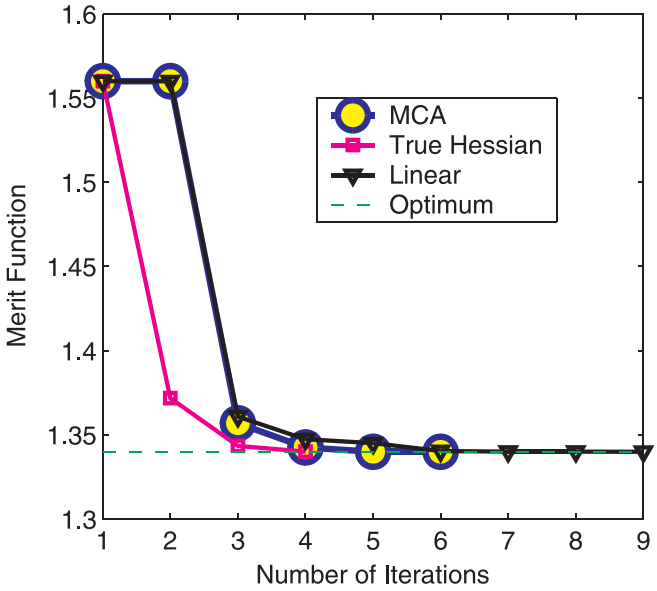

Fig. 3 Svanberg beam – reciprocal variables

erations for the second-order Taylor series. The MCA converged considerably quicker than SLP (Fig. 1). Use of the enriched reduced basis for MCA made insignificant changes to the results. Details of the iteration history are shown in Table 1. The negative trust ratio in iteration 6 indicates that the point was rejected.

When reciprocal variables were used for the constraint, the resulting sequential convex programming (SCP) approximate optimization converged smoothly, as long as the model trust region governed the move limits (Fig. 3). Impressively, the MCA without the aid of any intermediate variables converged as quickly as the SCP using reciprocal variables for the constraint. When the MCA was applied in the reciprocal variable space for the constraint, it accelerated the convergence, (6 iterations), comparable to the reciprocal, second-order Taylor series (4 iterations).

3.2 Large-scale cantilever beam optimization

Vanderplaats (1984) defined a similar cantilever beam design problem. The design variables are the height and width of the solid rectangular cross-sections of the beam. The objective is to minimize the volume subject to constraints on tip deflection (2.5 cm), bending stress and an aspect ratio constraint that the height to width ratio cannot exceed five. Thus, each beam segment has two design variables contributing to a structural response with nonseparable variables. The explicit functions are easily programmed for an arbitrary number beam segments, making this an ideal problem to test MCA as the num-

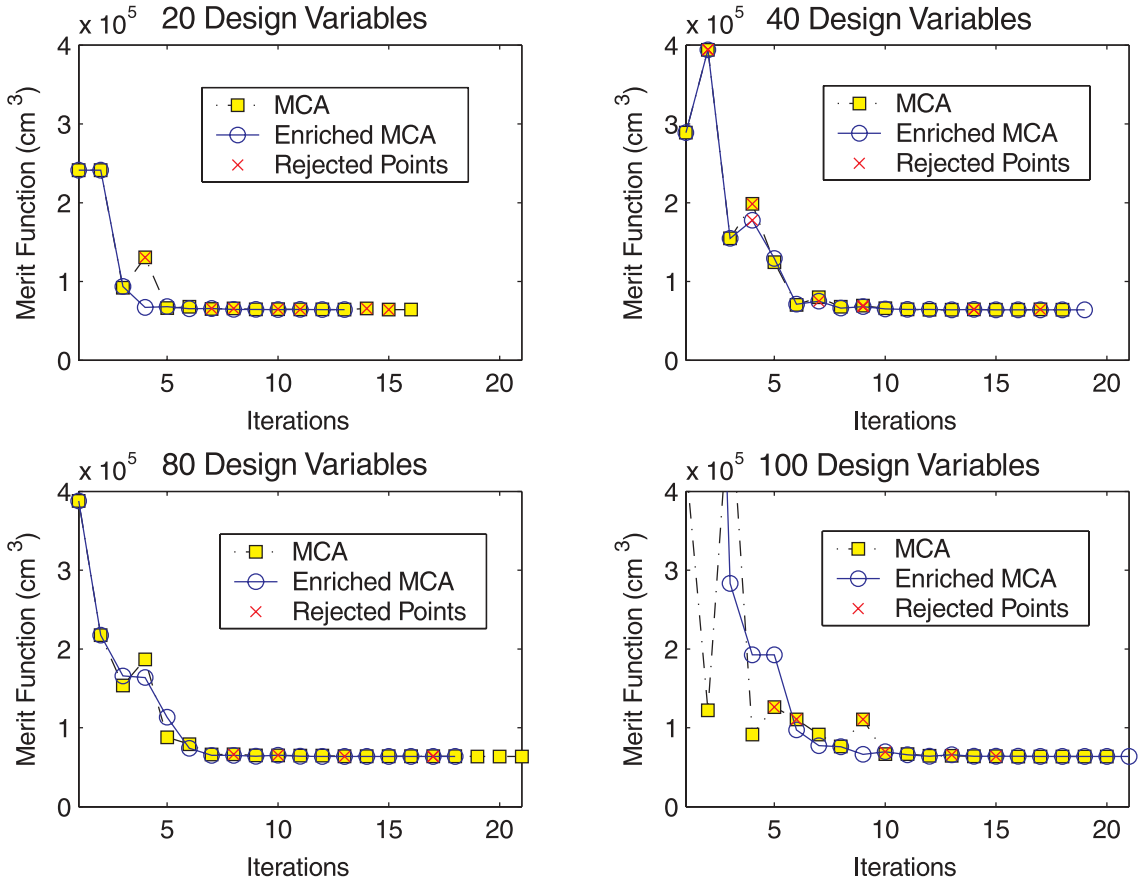

Fig. 4 Vanderplaats beam

Table 2 Enriched basis MCA for Vanderplaats beam

# DV	Iterations	Objective (cm^3)
10	7	65 418
20	13	64 244
40	19	63 842
80	18	63 757
100	21	63 704

ber of design variables increases. Although appropriate intermediate variables could be used for this example, it is important to emphasize that none were used here. Results for increasing number of beam segments are shown in Fig. 4. The enriched reduced basis was used in this example. Its primary effect was to increase the accuracy of the MCA in the early iterations, so that fewer points were rejected. However, the total number of iterations to converge was about the same. More importantly, the number of iterations to converge did not grow significantly as the number of designed beam segments increased. Table 2 shows the number of iterations to converge to within 0.1% of the optimum objective function value and 0.01% of the constraint, as well as the final value of the objective, using the enriched basis. The results bode well for the potential efficiency of MCA for large numbers of design variables, presuming an efficient algorithm is used to solve the approximate sub-problem.

3.3 Mixed-norm control synthesis

Another example (Fig. 5), the numerical solution of a mixed-norm, optimal control design problem (Doyle and Stein 1979) was selected to test MCA on a non-structural optimization problem for which good intermediate variables $\mathbf{y}(\mathbf{x})$ and responses $\mathbf{q}(\mathbf{y})$ are not known. Unlike the first beam example, which has separable design variables, this example is highly coupled.

The linear time-invariant model of the system dynamics is defined by P. The compensator to be designed is represented by K. Inputs to the system are zero-mean, unit intensity white Gaussian noise shown as w, and a bounded, but unknown, energy, input d. The measured values coming out of the system are shown as y and are input to the compensator. The control signal generated by the compensator, labelled u, is fed back to the system. The controlled outputs of the open-loop system are represented as z and e.

The H_2 and H_∞ problems have the following state space representations

$$\dot{x}_2 = A_2 x_2 + B_w w + B_2 u\,,$$

$$z = C_z x_2 + D_{zw} w + D_{zu} u\,,$$

$$y_2 = C_2 x_2 + D_{yw} w + D_{yu} u\,,$$

$$\dot{x}_\infty = A_\infty x_\infty + B_d d + B_\infty u\,,$$

$$e = C_e x_\infty + D_{ed} d + D_{eu} u\,,$$

$$y_\infty = C_\infty x_\infty + D_{yd} d + D_{yu} u\,.$$

The control synthesis problem is to find the compensator $K = (A_c, B_c, C_c)$ with the state space representation

$$\dot{x}_c = A_c x_c + B_c y$$

$$u = C_c x_c$$

which minimizes the H_2 norm of the transfer function T_{zw} from the noisy input w to the controlled output z, while

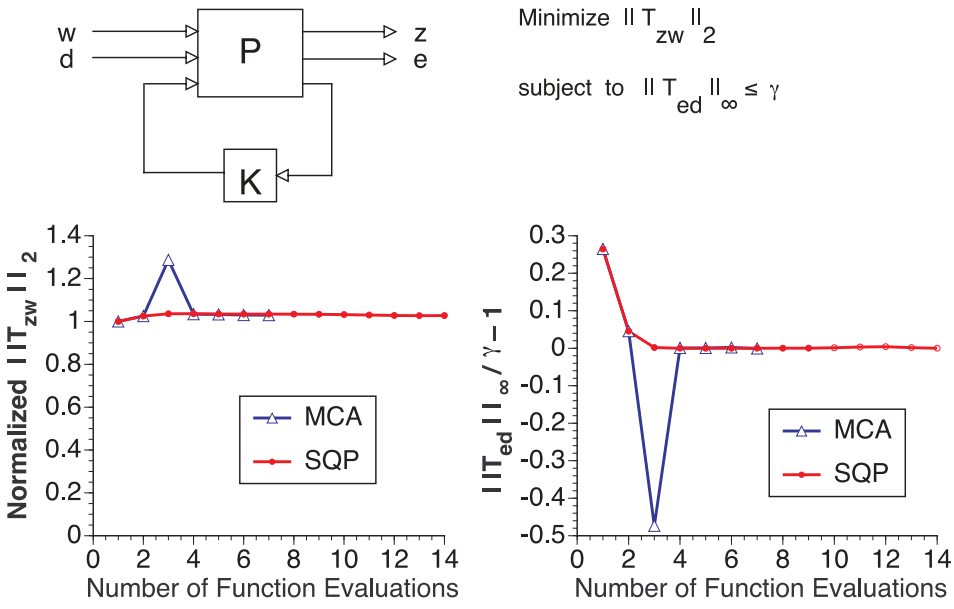

Fig. 5 Mixed-norm (H_2/H_∞) control synthesis

constraining the H_∞ norm of the transfer function T_{ed} from disturbance d to controlled output e below a specified level γ:

$$\min_{K_{adm}} ||T_{zw}||_2 \text{ subject to } ||T_{ed}||_\infty \leq \gamma$$

where the set of admissible compensators K_{adm} is the set of all stabilizing compensators of given fixed order. Hence, the design variables are the independent terms in the controller matrices A_c, B_c, and C_c. For more details see (Jacques *et al.* 1996).

A single-input/single-output example was used to demonstrate how to recover robustness with observers. The system's state space matrices are

$$A_2 = \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} 2 & 1 \end{bmatrix},$$

$$D_{yu} = 0\,.$$

Process and measurement noise are added to the model as

$$B_w = \begin{bmatrix} 35 & 0 \\ -61 & 0 \end{bmatrix}, \quad D_{yw} = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

The H_2 problem is a standard LQG problem with the following weighting matrices

$$C_z = 4 \begin{bmatrix} 5\sqrt{7} & \sqrt{5} \\ 0 & 0 \end{bmatrix},$$

$$D_{zw} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$D_{zu} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The H_∞ problem is a weighted sensitivity minimization problem used to improve tracking performance. The resulting H_∞ matrices are

$$A_\infty = \begin{bmatrix} A_2 & 0 \\ C_2 & -0.7 \end{bmatrix},$$

$$B_d = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$B_\infty = \begin{bmatrix} B_2 \\ 0 \end{bmatrix}$$

$$C_e = \begin{bmatrix} C_2 & 11 \end{bmatrix}, \quad C_\infty = \begin{bmatrix} C_2 & 0 \end{bmatrix},$$

$$D_{ed} = 1\,, \quad D_{eu} = 0\,,$$
$$D_{yd} = 1\,, \quad D_{yu} = 0\,.$$

The mixed H_2/H_∞ controller was run at the order of the H_2 problem (second order), yielding seven independent design variables in the controller matrices. Thus, the design variables are the four terms in A_c, two terms in B_c, plus the single term C_c of the compensator K. A constraint was imposed to reduce the H_∞ norm from 15.18 at the H_2 optimum design to 12. The trust region was initiated with 25% move limits. Iteration histories for the objective normalized by the optimal H_2 norm (493.8) and the H_∞ constraint are be plotted in Fig. 5. The MCA converged in half as many iterations (7) as SQP (14).

3.4 Discussion

Successive approximate optimization converged more quickly with MCA than with single-point approximations for all three examples. The beam example in Sect. 3.1 demonstrated that MCA without intermediate variables could converge at least as quickly as the single-point approximation with reciprocal variables for the displacement constraint. The results presented in Figs. 2 and 3 were obtained using the same trust region strategy, in order to highlight only the effect of the differing approximations. Although this example involved separable design variables, MCA does not take advantage of that information until there are more previous points than design variables, because the coefficients of the purely cubic terms correspond to the reduced subspace in which the original design variables are coupled. The design variables in the next two examples were not separable.

The beam example in Sect. 3.2 demonstrated that the MCA scales well to problems with large numbers of design variables. A good measure of efficiency for sequential MCA is the comparison with sequential approximate optimization using the actual Hessian matrix in a second-order Taylor series approximation. Sequential MCA compared favourably in this regard. As expected, it never outperformed the second-order Taylor series approximation, since MCA requires several iterations to gather curvature information. Use of MCA for the structural optimization examples in Sects. 3.1 and 3.2 did not display dramatic improvements over the mature approximation methods used in modern structural optimization programs; however, the real potential is for multidisciplinary problems where appropriate intermediate variables and responses are unknown. This benefit was realized in the H_2/H_∞ optimal control problem in Sect. 3.3, where sequential MCA outperformed SQP. Furthermore, the MCA does not suffer from the curse of dimensionality as SQP and second-order Taylor series approximation do as the number of design variables increases.

The effectiveness of MCA was demonstrated within the framework of sequential approximate optimization using a trust region strategy. More generally, MCA can be considered a response surface technique. The author observed its accuracy to be similar to or better than the previously developed MQA, the accuracy of which was demonstrated for arbitrarily generated design points (Canfield 1994). Implementing MCA as a response surface technique requires similar choices as any other, such as, how many points to include and at what locations? In the preceding examples, when the number of previous points exceeded the number of design variables, n, only the closest n linearly independent design vectors were retained (except for the large-scale problem in Sect. 3.2 where the number of previous points was always fewer than n). The major advantage of MCA is that the number of available design points determines the number of higher-order terms, rather than the number of design variables dictating how many design points must be sampled.

Noisy numerical simulations affect the implementation of MCA as a response surface technique. By virtue of being a low-order polynomial, MCA will exhibit a smooth approximation when the design points are widely spaced. When an optimization process dictates the selection of design points, as in the preceding examples, widely spaced points are typical in the early iterations. As the spacing of design points becomes closer in the final stages of convergence, a low signal-to-noise ratio will destroy the accuracy of MCA, unless an over-determined system is created by retaining more previous points than the number of cubic terms. Also, the gradient appearing in the linear term may need smoothing. In the final iterations of an optimization-guided process, an ample amount of previous points makes an over-determined system easy to accommodate. In a design of experiments (DOE) setting, an over-determined MCA should be used for noisy simulations in the same manner as is done for conventional response surfaces. This simply means that more points than the number of cubic terms should be generated for the DOE. Then a pseudo-inverse should be used in (22) and (25).

4 Conclusion

A multipoint cubic approximation was developed for numerical optimization. Its derivation hinged on the feature that previously known function and gradient values are recovered by the approximation. Reducing the quadratic and cubic terms to a subspace of the design variables spanned by the previous design points made for much more efficient calculation and subsequent use of the approximate reduced Hessian and cubic terms. Numerical examples demonstrated efficient global convergence.

In contrast to the Hessian updates used in quasi-Newton methods, the multipoint cubic approximation is not restricted to being positive definite, or even symmetric, although it proved advantageous in numerical optimization to use only positive quadratic/cubic corrections for constraints. Furthermore, its "memory" of past design points is based on both function values and gradients, whereas the quasi-Newton updates use only gradients. Independent use of MCA for the objective function and active constraints is partly responsible for the quicker convergence.

For structural optimization problems, convergence for MCA without intermediate variables was comparable to convergence achieved for problems that benefit from reciprocal variables. The MCA offers the greatest benefit when appropriate intermediate variables or responses are not known, as the H_2/H_∞ control design problem demonstrated. Moreover, the reduced basis for the quadratic and cubic terms avoids the "curse of dimensionality" for problems with many design variables. The subspace for the higher-order terms does not reduce the search space, however, because the linear terms are not reduced. Very large problems, though, will require efficient solution of the approximate sub-problem, which dominated the computational time for the first two examples, which had explicit functions. Design problems with computationally intensive simulations for function and gradient evaluations are the prime candidates to benefit from the efficiency of MCA models. Future work should exploit the potential for parallel computation, when MCA is applied to problems with many constraints.

Acknowledgements The Air Force Office of Scientific Research funded this research. The author gratefully acknowledges the support of the computational mathematics program manager, Maj. William Hilbun.

References

Alexandrov, N.M.; Dennis, J.E., Jr.; Lewis, R.M.; Torczan, V. 1998: A trust-region framework for managing the use of approximation models in optimization. *Struct Optim* **15**(1), 16–23

Barthelemy, J.M.; Haftka, R.T. 1993: Approximation concepts for optimum structural design – a review. *Struct Optim* **5**, 129–144

Branch, M.A.; Andrew, G. 1996: *Optimization Toolbox User's Guide The MathWorks*. Natick, Mass.

Canfield, R.A. 1994: Multipoint quadratic approximation for numerical optimization. In *5th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Panama City, Florida, September 7–9 1994

Canfield, R.A. 2001: Sequential multipoint quadratic approximation for numerical optimization. In *42nd AIAA/ASME/-ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Seattle, Washington, April 16–19 2001

Conn, A.R.; Gould, N.I.M.; Toint, P.L. 1991: A globally convergent augmented lagrangian algorithm for optimization

with general constraints and simple bounds. *SIAM J Numer Anal* **28**(2), 545–572

Doyle, J.C.; Stein, G. 1979: Robustness with observers. *IEEE Trans Automatic Control* **AC-24** (4), 607–611

Fadel, G.M.; Riley, M.F.; Barthelemy, J.-F.M. 1990: Two point exponential approximation method for structural optimization. *Struct Optim* **2**, 117–124

Fletcher, R. 1981: *Practical Methods of Optimization*. New York: John Wiley

Gill, P.E.; Murray, W.; Wright, M.H. 1981: *Practical Optimization*. New York: Academic Press

Haftka, R.T.; Gurdal, Z. 1992: *Elements of Structural Optimization*. Kluwer Academic Publishers

Haftka, R.T.; Watson, J.A.; Nachlas, L.T.; Rizzo, T.; Desai, R. 1987: Two-point constraint approximation in structural optimization. *Comput Methods Appl Mech Eng* **60**, 289–301

Jacques, D.R.; Canfield, R.A.; Ridgely, D.B.; Spillman, M.S. 1996: A MATLAB toolbox for fixed order, mixed-norm control synthesis. *IEEE Control Syst* **15**(5), 26–44

Miura, H.; Schmit, L.A., Jr. 1978: Second order approximation of natural frequency constraints in structural synthesis. *Int J Numer Methods Eng* **13**(2), 336–351

Rasmussen, J. 1990: Accumulated approximation – a new method for structural optimization by iterative improvement. In *3rd AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, San Francisco, CA

Schmit, L.A., Jr.; Miura, H. 1977: Approximation concepts for efficient structural synthesis. Technical Report 2552, NASA, March 1977

Snyman, J.A.; Stander, N. 1994: New successive approximation method for optimum structural design. *AIAA J* **32**(6), 1310–1315

Svanberg, K. 1987: The method of moving asymptotes-a new method for structural optimization. *Int J Numer Methods Eng* **24**, 359–373

Toropov, V.V.; Filatov, A.A.; Polynkin, A.A. 1993: Multiparameter structural optimization using FEM and multipoint explicit approximations. *Struct Optim* **6**, 7–14

Vanderplaats, G.N. 1984: *Numerical Optimization Techniques for Engineering Design: With Applications*. New York: McGraw-Hill

Wujek, B.A.; Renaud, J.E. 1998: New adaptive move-limit management strategy for approximate optimization, part i. *AIAA J* **36**(10), 1911–1921

Xu, S.; Grandhi, R.V. 1998: Effective two-point function approximation for design optimization. *AIAA J* **36**(12), 2269–2275

Xu, S.; Grandhi, R.V. 2000: Multipoint approximation development: Thermal structural optimization case study. *Int J Numer Methods Eng* **48**, 1151–1164