INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2009; **61**:591–605
Published online 23 December 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1967

Multi-objective turbomachinery optimization using a gradient-enhanced multi-layer perceptron

M. C. Duta[1,*,†] and M. D. Duta[2]

[1]*Oxford E-Science Research Centre, University of Oxford, Oxford OX1 3QG, U.K.*
[2]*Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, U.K.*

SUMMARY

Response surface models (RSMs) have found widespread use to reduce the overall computational cost of turbomachinery blading design optimization. Recent developments have seen the successful use of gradient information alongside sampled response values in building accurate response surfaces. This paper describes the use of gradients to enhance the performance of the RSM provided by a multi-layer perceptron. Gradient information is included in the perceptron by modifying the error function such that the perceptron is trained to fit the gradients as well as the response values. As a consequence, the back-propagation scheme that assists the training is also changed. The paper formulates the gradient-enhanced multi-layer perceptron using algebraic notation, with an emphasis on the ease of use and efficiency of computer code implementation. To illustrate the benefit of using gradient information, the enhanced neural network model is used in a multi-objective transonic fan blade optimization exercise of engineering relevance. Copyright © 2008 John Wiley & Sons, Ltd.

Received 30 April 2007; Revised 15 July 2008; Accepted 10 October 2008

KEY WORDS: artificial neural networks; multi-layer perceptron; gradient-enhanced response surface; adjoint method; turbomachinery optimization

1. INTRODUCTION

Global design optimization of turbomachinery blading typically requires a large number of evaluations of one or several objective functionals, which are normally based on expensive Navier–Stokes flow calculations. To alleviate this cost penalty, state-of-the-art global optimization routinely uses response surface models (RSMs) that approximate the variation of the objectives in the design space [1]. RSMs are derived from a limited number of exact objective values (samples) obtained

*Correspondence to: M. C. Duta, Oxford E-Science Research Centre, University of Oxford, Oxford OX1 3QG, U.K.
†E-mail: mihai.duta@oerc.ox.ac.uk

Contract/grant sponsor: U.K. Department for Trade and Industry and Rolls-Royce plc

from expensive direct flow simulations. Response surface methodology is diverse and ranges from local definitions valid within a limited region, e.g. low-order polynomial approximations, to global models, e.g. kriging and artificial neural networks (ANNs), [2]. Also, there exist several approaches to updating the RSMs with new exact objective evaluations during the course of the optimization process [3].

RSMs are conventionally based only on the sampled objective values, but recent research has also explored the use of objective gradient information (inexpensively available from the adjoint method) alongside the response values [4]. This idea has been tried in conjunction with both kriging [5] and ANNs [6]. The expected advantage of a gradient-enhanced RSM is a higher accuracy of the response surface, especially noticeable in higher-dimensional design spaces across gaps in the sampled objective values.

Owing to their simplicity and flexibility, ANNs represent a class of RMSs that have been receiving increasing attention in the context of blade optimization [7]. Gradients have been used by ANNs as a regularization mechanism [8], which ensures a degree of smoothness for the output. Recently, Giannakoglou *et al.* [6] has formulated a general framework of gradient-enhanced ANNs and demonstrated its usefulness on several inverse-design uni-modal optimization problems.

Building on similar ideas, this paper re-formulates the gradient enhancement of a multi-layer perceptron (MLP) in the simplest terms, using algebraic notation. This has the advantage of being both easy to follow and leading to an optimally fast computer implementation using Basic Linear Algebra Subprograms (BLAS), the high-performance computing standard for matrix and vector operations. The paper also discusses the relevant aspects of MLP training process as well as the robustness of the gradient-enhanced scheme.

The paper expands the application base of the gradient-enhanced MLP to include exploratory design optimizations, with an objective response characterized by a complex, multi-modal landscape. A realistic multi-objective turbomachinery optimization exercise of engineering relevance is presented to demonstrate the use of the method in practical terms.

2. CONVENTIONAL MLPs

2.1. Definitions

ANNs are nonlinear models that map from an input to an output space in multi-dimensions. An important type of ANN is the MLP, a feed-forward network with several layers of nonlinear units. According to the Cybenko theorem, the two-layer MLP is a universal approximator, i.e. it can approximate any continuous real input/output mapping with arbitrary accuracy, given a sufficiently large number of units, [8]. This work is concerned exclusively with this type of MLP, simply called the MLP throughout the rest of the paper.

Figure 1 gives a schematic representation of the MLP. Each unit in the hidden layer takes as input a biased weighted sum of the MLP inputs and passes it through a nonlinear activation function, which is differentiable with respect to the weights. Then, each unit in the output layer takes as input a biased weighted sum of the outcome of the hidden layer and passes this to the MLP output.

The values of the weights and biases are found by a training process, during which the MLP model is adjusted to give a best fit to the target training data.

DOI: 10.1002/fld

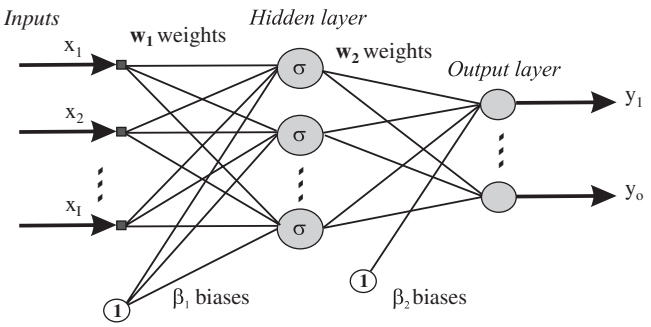

Figure 1. Schematic representation of a MLP with I inputs and O outputs.

In a design optimization context, the input to the MLP is a vector of design parameter values and the outputs are the objective functionals used as figures of merit during optimization. The data on which the MLP is trained are furnished by the design of experiment (DoE), which samples the design space at a number of points, in a statistically meaningful manner.

2.2. Formulation

Mathematically, an MLP with I inputs, H units in the hidden layer and O outputs is described by

$$\begin{aligned} \mathbf{u} &= \mathbf{w}_1 \cdot \mathbf{x} + \beta_1 \\ \mathbf{v} &= \sigma(\mathbf{u}) \\ \mathbf{y} &= \mathbf{w}_2 \cdot \mathbf{v} + \beta_2 \end{aligned} \tag{1}$$

in which $\mathbf{x}$ is the vector of I inputs, $\mathbf{y}$ the vector of O outputs and $\cdot$ denotes matrix multiplication. The input weights are represented by the $H \times I$ matrix $\mathbf{w}_1$ and the biases by the H-dimensional vector β_1. Similarly, the output weights are given by the $O \times H$ matrix $\mathbf{w}_2$ and the biases by the O-dimensional vector β_2. The weights and biases control the strength of the connections between units.

$\mathbf{u}$ and $\mathbf{v}$ are the respective input and output of the hidden layer and are both H-dimensional vectors. The nonlinear activation function σ maps each entry in the input $\mathbf{u}$ to the corresponding entry in $\mathbf{v}$. The purpose of this function is to adjust the contribution of the input units to the output result. There are many functions suited to this purpose; the hyperbolic tangent function $\sigma(\mathbf{u})=\tanh(\mathbf{u})$ is used here.

The weights and biases of the MLP are computed in an iterative training process during which the values are adjusted to best fit the DoE data. Thus, the DoE consists of a number M of points $\mathbf{x}_m$ in the design space and the associated response values $\bar{\mathbf{y}}_m$, $m=1,\ldots,M$, computed at those points through direct flow simulations. The degree at which the MLP output $\mathbf{y}_m=\mathbf{y}(\mathbf{x}_m)$ at the DoE points fit the target response $\bar{\mathbf{y}}_m$ is very commonly defined in the least-squares sense and quantified by the error function

$$e = \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{y}_m - \bar{\mathbf{y}}_m\|^2 \tag{2}$$

which is minimized to achieve the best fit.

DOI: 10.1002/fld

The definition of the MLP in (1) can be carried one step further towards vectorization. This is achieved by collapsing the operations pertaining to different points in the design space into single matrix operations. For instance, the MLP output for all the DoE points $\mathbf{x}_m$ can be carried out simultaneously by allowing $\mathbf{x}$ to include all the points, i.e. $\mathbf{x}$ becoming an $I \times M$ matrix with the columns storing the DoE points. Thus, the vectorized MLP definition is

$$\begin{aligned} \mathbf{u} &= \mathbf{w}_1 \cdot \mathbf{x} + \beta_1 \cdot \mathbf{1}^{\mathrm{T}} \\ \mathbf{v} &= \sigma(\mathbf{u}) \\ \mathbf{y} &= \mathbf{w}_2 \cdot \mathbf{v} + \beta_2 \cdot \mathbf{1}^{\mathrm{T}} \end{aligned} \tag{3}$$

where $\mathbf{1}$ is an M-dimensional vector of unit values. From here on, $\mathbf{x}$ signifies the $I \times M$ matrix of DoE points and $\mathbf{y}$ is the $O \times M$ matrix of MLP outputs at those points. Evidently, the columns of $\mathbf{v}$ and $\mathbf{u}$ contain the hidden layer inputs and outputs corresponding to all the M DoE points.

The meaning of M in (3) depends in fact on the situation. While M is indeed the number of DoE points during the training phase, it can, for instance, become the number of members in a Genetic Algorithm (GA) population during the use of the MLP as a RSM, with $\mathbf{x}$ representing M population points that are not in the DoE.

The definition (3) as well as the following vectorized gradient-enhanced formulation are of notable practical importance, as the computer code implementation should make use of a fast BLAS implementation of the matrix and vector operations. This has a beneficial impact on code performance, in both the training phase and in using the MLP as a RSM, with the operations carried out at optimal speed.

Lastly, the least-squares function changes with the vectorized notations above to

$$e = \tfrac{1}{2} \|\mathbf{y} - \bar{\mathbf{y}}\|_{\mathrm{F}}^2 \tag{4}$$

where $\bar{\mathbf{y}}$ is the $O \times M$ matrix of DoE samples and $\|\mathbf{y}\|_{\mathrm{F}}$ is the Frobenius norm of $\mathbf{y}$.

2.3. Training

Training the MLP means finding the weights and biases in (3) that minimize the least-squares error function (4). The minimization of e constitutes an unconstrained optimisation problem in its own right for which there exist multiple solutions [8]. Because the error function is normally highly multi-modal (it has multiple local optima), these solutions include global-search algorithms like GAs and simulated annealing (SA). However, the most popular by far are two iterative gradient-based methods; the *Levenberg–Marquardt* (LM), a minimizer for sums of nonlinear squares, combining features of the Newton iteration with those of the gradient descent, and the *Scaled Conjugate Gradients* (SCG), a minimizer belonging to the class of conjugate gradient algorithms.

The popularity of gradient-based methods is due to the inexpensive availability of the gradients of the error function e in the weight-bias space, obtained though the *Back-propagation* algorithm. Back-propagation is in fact the *adjoint* (backward) propagation of the error gradients through the MLP, a description more familiar to the optimization community. Its attraction is that gradients

DOI: 10.1002/fld

are obtained at a cost linear with the dimension of the weight-bias space. The back-propagation process is described by

$$\begin{aligned} \frac{\partial e}{\partial \mathbf{w}_2} &= \delta_2 \cdot \mathbf{v}^{\mathrm{T}}, \quad \frac{\partial e}{\partial \beta_2} = \delta_2 \cdot \mathbf{1}^{\mathrm{T}} \\ \frac{\partial e}{\partial \mathbf{w}_1} &= \delta_1 \cdot \mathbf{x}^{\mathrm{T}}, \quad \frac{\partial e}{\partial \beta_1} = \delta_1 \cdot \mathbf{1}^{\mathrm{T}} \end{aligned} \tag{5}$$

This algorithm propagates the error gradients δ_2 in a backward feed, from the MLP output through the hidden layer gradient δ_1 to the MLP input. During this process, all the gradients with respect to the weights and biases are computed. The intermediary gradients are given by

$$\begin{aligned} \delta_2 &= \frac{\partial e}{\partial y} = \mathbf{y} - \bar{\mathbf{y}} \\ \delta_1 &= \sigma'(\mathbf{u}) \odot (\mathbf{w}_2^{\mathrm{T}} \cdot \delta_2) \end{aligned} \tag{6}$$

the symbol $\odot$ representing the entry-by-entry multiplication of two matrices of same dimension.

By their local scope, gradient-based methods tend to be trapped within regions of local optima of the multi-modal landscape of the error function e. To avoid this shortcoming while still retaining the speed of the gradient-based minimization, a restarted approach is traditionally followed. In this approach, a number of local minimizations are started independently from different points in the weight-bias space, the global optimum taken to be the best found across the different minimizations. In theory, restarted optimizations find the global optimum with a probability that tends to 1 asymptotically with the number of restarts [9]. In practice, a small number of restarts are sufficient, which can be determined on a problem basis. As the independent gradient-based minimizations can be trivially parallelized, the restarted approach is a very practical combination of speed and simplicity.

While the number of inputs and outputs is set by the problem specifications, the number of hidden layer units is not and finding a good value is part of building the MLP. This is normally achieved by training the MLP for a range of hidden layer sizes and identifying the number after which the benefit of including more hidden units is practically lost. This was done for instance for the transonic rotor in Section 5; the variation of the trained error with the number of hidden layer units in Figure 5(a) indicates that the optimum number is $H = 10$, meaning that including more does not further reduce the training error significantly.

Another practical aspect is the number of iterations in the MLP training; although training to full convergence of the error function e to the global minimum sounds reasonable, this is not done for two reasons. First, the convergence rate becomes impractically slow after a relatively small number of iterations. Second, while the fully converged MLP fits the DoE data better than a partially converged one, it may be a worse overall interpolator, its predictions overshooting in regions between the DoE points [8]. A common solution to finding a good compromise between the accuracy of fitting and good interpolation properties is to split the DoE data into two sets; the larger one is used for training in the normal fashion and the smaller is kept for validating the predictions made by the MLP trained on the first set. Figure 5(b), which depicts one MLP error convergence history on the transonic rotor DoE, is typical; the error measured on the training set is constantly decreasing while the error on the validation set starts to increase slightly after a certain number of iterations, at which point the training is stopped.

DOI: 10.1002/fld

3. GRADIENT-ENHANCED MLPs

3.1. Definition

To enhance any RSM with gradient information, it is assumed that the gradients of the objective functions are obtained at the DoE points $\mathbf{x}_m$ alongside the objective response values themselves. Then, both the forward and the backward propagation schemes (3) and (5) are modified to include gradients alongside the response.

First, the forward propagation is changed to map the input $\mathbf{x}$ to the gradients of $\mathbf{y}$ as well as to the response $\mathbf{y}$ itself:

$$\begin{aligned}
&\mathbf{u}=\mathbf{w}_1\cdot\mathbf{x}+\beta_1\cdot\mathbf{1}^{\mathrm{T}}, \quad \frac{\partial\mathbf{u}}{\partial x_i}=\mathbf{w}_1\cdot\frac{\partial\mathbf{x}}{\partial x_i}\\
&\mathbf{v}=\sigma(\mathbf{u}), \quad \frac{\partial\mathbf{v}}{\partial x_i}=\sigma'(\mathbf{u})\odot\frac{\partial\mathbf{u}}{\partial x_i}\\
&\mathbf{y}=\mathbf{w}_2\cdot\mathbf{v}+\beta_2\cdot\mathbf{1}^{\mathrm{T}}, \quad \frac{\partial\mathbf{y}}{\partial x_i}=\mathbf{w}_2\cdot\frac{\partial\mathbf{v}}{\partial x_i}
\end{aligned} \tag{7}$$

While the equations in the left column are identical to those in (3), the equations in the right column are the linearization of the operations in the left and describe the forward propagation of gradients. Thus, the gradient $\partial\mathbf{y}_i=\partial\mathbf{y}/\partial x_i$ with respect to the ith variable x_i is obtained by propagating the gradient matrix $\partial\mathbf{x}/\partial x_i$ (with zero entries except its ith row of unit values) through the MLP using the original weights of (3). The MLP topology does not change but the gradients are output alongside the response $\mathbf{y}$.

Second, the gradients are included in the MLP training process by modifying the least-squares error function to become

$$e=\frac{1}{2}\|\mathbf{y}-\bar{\mathbf{y}}\|_{\mathrm{F}}^2+\frac{1}{2}\sum_{i=1}^{I}\|\partial\mathbf{y}_i-\partial\bar{\mathbf{y}}_i\|_{\mathrm{F}}^2 \tag{8}$$

Thus, minimization of e ensures that the MLP output $\mathbf{y}$ fits the DoE response data $\bar{\mathbf{y}}$ and, at the same time, the MLP gradients $\partial\mathbf{y}_i$ fit the DoE gradients $\partial\bar{\mathbf{y}}_i$, for all i inputs. Notice that the ith variable gradients $\partial\mathbf{y}_i$ and $\partial\bar{\mathbf{y}}_i$ are both matrices of size $O\times M$. Effectively, they could be described using third-order tensor notation but that would be against the simple algebra/fast implementation philosophy.

With the above changes in forward propagation and error function, the back-propagation of the gradient-enhanced MLP is:

$$\begin{aligned}
&\frac{\partial e}{\partial\mathbf{w}_2}=\delta_2\cdot\mathbf{v}^{\mathrm{T}}+\sum_{i=1}^{I}\partial\delta_{2,i}\cdot\left(\frac{\partial\mathbf{v}}{\partial x_i}\right)^{\mathrm{T}}, \quad \frac{\partial e}{\partial\beta_2}=\delta_2\cdot\mathbf{1}^{\mathrm{T}}\\
&\frac{\partial e}{\partial\mathbf{w}_1}=\delta_1\cdot\mathbf{x}^{\mathrm{T}}+\sum_{i=1}^{I}\left(\partial\delta_{1,i}^1\cdot\mathbf{x}^{\mathrm{T}}+\partial\delta_{1,i}^2\cdot\left(\frac{\partial\mathbf{x}}{\partial x_i}\right)^{\mathrm{T}}\right), \quad \frac{\partial e}{\partial\beta_1}=\delta_1\cdot\mathbf{1}^{\mathrm{T}}+\sum_{i=1}^{I}\partial\delta_{1,i}^1\cdot\mathbf{1}^{\mathrm{T}}
\end{aligned} \tag{9}$$

DOI: 10.1002/fld

While the gradients δ_1 and δ_2 carry the same meaning as in (5), the other back-propagated intermediate gradient quantities are defined by:

$$\partial\delta_{2,i} = \partial\mathbf{y}_i - \partial\bar{\mathbf{y}}_i$$

$$\partial\delta^1_{1,i} = \sigma''(\mathbf{u}) \odot \left(\frac{\partial\mathbf{u}}{\partial x_i}\right) \odot (\mathbf{w}_2^{\mathrm{T}} \cdot \partial\delta_{2,i}) \tag{10}$$

$$\partial\delta^2_{1,i} = \sigma'(\mathbf{u}) \odot (\mathbf{w}_2^{\mathrm{T}} \cdot \partial\delta_{2,i}) \tag{11}$$

3.2. Training

The training of the gradient-enhanced MLP is similar to that of the conventional MLP. The optimal number of hidden units has to be evaluated on a problem-based case and the number of iterations used for training has to be established, as described in Section 2.3.

An interesting observation is that while the training of the conventional MLP can be also achieved using the LM algorithm, this becomes impractical in the case of the gradient-enhanced MLP. The algorithm is specialized in the minimization of sums of squares of nonlinear functions and requires the availability of the Hessian of the sum. For the gradient-enhanced MLP error function, this would lead to third-order tensors and, consequently, to a cumbersome formulation and an error-prone implementation. The performant SCG, which uses the gradients of the error function directly, is simpler to use and was preferred throughout this work.

4. VALIDATION

Before employing the gradient-enhanced MLP in optimization, the correctness and robustness of the model are validated by three tests using simple analytic functions.

First, the Rastrigin function defined as $\sum_{i=1}^{I}(x_i^2 - 10\cos(2\pi x_i) + 10)$ in I dimensions is employed to highlight the usefulness of the gradient information. Figure 2(a) shows contours of the

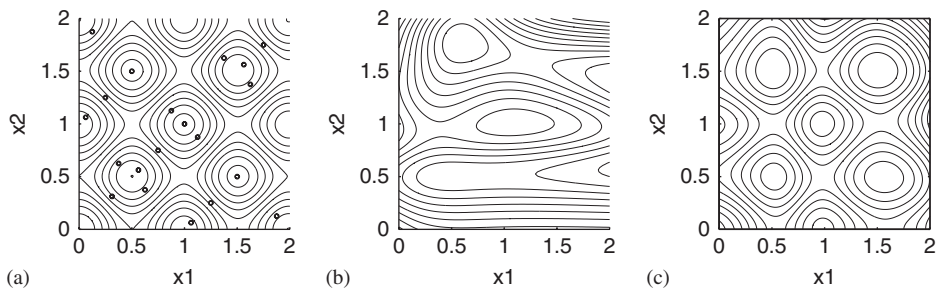

Figure 2. Multi-modal variation of the Rastrigin function (a). Re-construction using a conventional MLP (b) and 20 samples at the locations shown in black marks in (a). Re-construction using a gradient-enhanced MLP (c) and half of the samples shown in (a).

DOI: 10.1002/fld

multi-modal variation of this function in two dimensions, with both variables ranging from 0 to 2. Twenty sampling points computed by the LP_τ DoE [10] are also shown; this is a number typical for the first sampling of a 2D design space without prior knowledge of the response. Figure 2(b) depicts the contours of the conventional MLP built using the 20 samples and shows that the response characteristics are not well captured. On the contrary, Figure 2(c) shows the contours of the gradient-enhanced MLP, with a clearly correct re-construction of the original function from the DoE. Moreover, the number of samples used to build the enhanced MLP is 10 – the first half of the DoE used by the conventional MLP! This is important because in a design optimization the cost of the computing the adjoint gradients (in any number of dimensions!) is roughly the same as the cost of obtaining the response values, using half the number of samples ensures the cost of the two re-constructions is the same.

Even in the case of a unimodal design landscape, the use of gradient information helps prevent undesirable oscillations of the RSM due to the sparsity of experiments and thus facilitates the training of the MLP. Also, the usefulness of the enhanced MLPs likely to increase in higher-dimensional design spaces, where the adjoint method computes the gradients at no higher cost than in the 2D case, whereas the conventional MLP requires additional expensive flow simulations to obtain more information.

The last two tests involve a function whose variation along one variable is depicted in Figure 3(a); its definition in an multi-dimensional space is given simply by the summation of these variations in each variable. This test function is characterized by a multi-modal shape with a known global maximum. The second test is carried out on the function thus defined in four dimensions and measures the variation of an error measure with the size of the DoE, Figure 3(b). The number of points at which the function is sampled is varied and the DoE is used to train two MLPs with a fixed number of hidden layer units. The error measure is the root-mean-square difference between the MLP output and true function values computed at a number of new, unsampled points. The error of the gradient-enhanced MLP reduces faster with the size of the DoE than that of the conventional MLP; the same error level is achieved by the gradient-enhanced MLP for less than half the size of the DoE. Moreover, the stall noticeable in the error variation, which is due to fixing the number of hidden layers, takes place earlier and at a higher level for the conventional MLP than for the gradient enhanced.

The third test performs two independent optimizations using a GA [11] assisted by the respective conventional and gradient-enhanced MLP trained off-line. The same function in four dimensions as for the previous test is sampled and in both cases, the MLP is trained and the GA performs an unconstrained maximization using the MLP as a RSM. The optimal point thus found is incorporated into the DoE using the exact function value and the training-maximization step is repeated. Using the Euclidian distance between the update and the known global maximum as an error measure, Figure 3(c) depicts the convergence of the optimization with the DoE update iteration. The gradient-enhanced MLP converges faster to the global maximum than the conventional and, for a fixed level of accuracy, it requires fewer DoE updates.

5. TRANSONIC FAN OPTIMIZATION

To illustrate the use of the gradient-enhanced MLP, two design exercises have been carried out on the Nasa Rotor 37 public domain testcase. This is a highly loaded transonic compressor, designed

DOI: 10.1002/fld

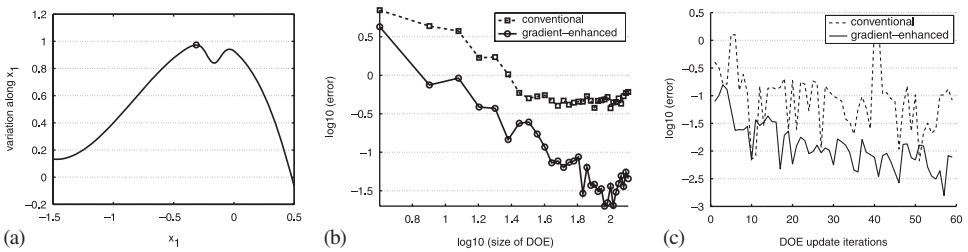

Figure 3. Variation of a multi-modal test function along one variable (a); the global maximum is circled. Variation of the MLP error measured at a series of unsampled points with the size of the DoE (b), showing an increased accuracy of the gradient-enhanced MLP. Convergence of two GA optimizations based on MLP approximations (c), showing superior convergence properties of the gradient-enhanced MLP.

and studied experimentally at the NASA Glenn Research Center [12], which contains many features of a real-life engineering design case. The optimization software used was Soft [11], an optimization environment with state-of-the-art libraries for local and global search. The geometry parametrization and meshing was achieved by the multi-block meshing system Padram [13]. The flow solutions, response and gradients were computed using the turbomachinery flow solver Hydra [14], using the Spalart–Allmaras turbulence model for the viscous flow and multiple grids for the acceleration of solution convergence.

At any given experiment point, the response and its gradients in the design space are obtained using the adjoint method [15]. Thus, the gradients of a single objective in the design space are efficiently computed at the cost of a single adjoint Navier–Stokes solution, which is independent of the dimension of this space. The adjoint code is obtained from the original nonlinear flow solver with the aid of Automatic Differentiation software [16]. For practical purposes, the cost of the adjoint solution is equal to that of the direct flow calculation involved in obtaining the objective.

The first design exercise is a simple unconstrained maximization of the stage isentropic efficiency in a low-dimensional design space. The sweep of the blade is the main re-design parameter as this is known to have a significant influence on the shock loss and tip entropy generation [17]. Hence, the first design parameter is the sweep (both forward and backward) of the blade tip along its chord. The second parameter is the blade control height from which the sweep re-design has an effect; the blade is swept only from that height towards the tip. The radial sweep distribution thus follows a spline variation from 0 at the specified height to the specified maximum sweep at the tip.

This 2D design space was sampled at 64 points defined by the LP_τ DoE and the isentropic efficiency and its gradients with respect to the design parameters were computed. The calculations were carried out at a fixed back-pressure at the rotor working condition (97.91% of choke massflow, 100% engine speed). A small number of the DoE points did not produce data because of failed mesh generation or flow calculation and were discarded. The gradient-enhanced MLP was trained using the sampled response and gradients and Figure 4 depicts the contours of its output over a range of control height and sweep. The variation of efficiency is smooth and has a maximum at 78% control height and 14% of chord forward sweep. The sample points are shown as black dots and the optimum is indicated by a cross. Owing to the smooth variation and the relatively large size of the DoE, the optimum predicted by the MLP is within 1% off the direct simulated value.

DOI: 10.1002/fld

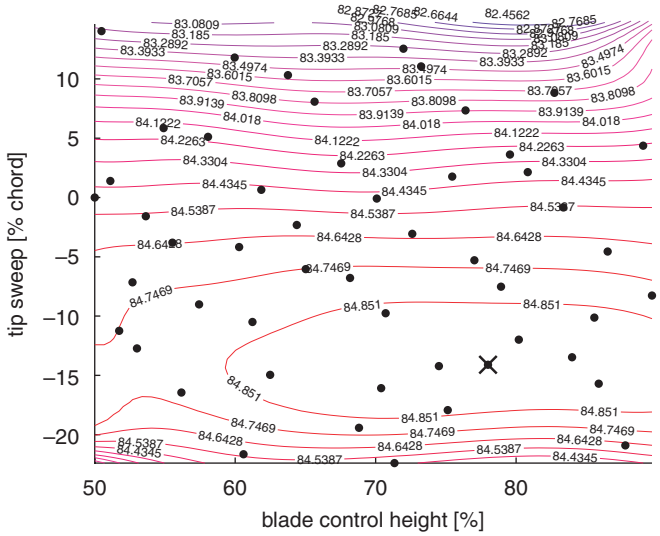

Figure 4. Smooth variation of efficiency in the 2D tip sweep/control height design space. The sampled values are shown as black dots. The optimum is characterized by a forward sweep of the blade tip and is marked with a cross.

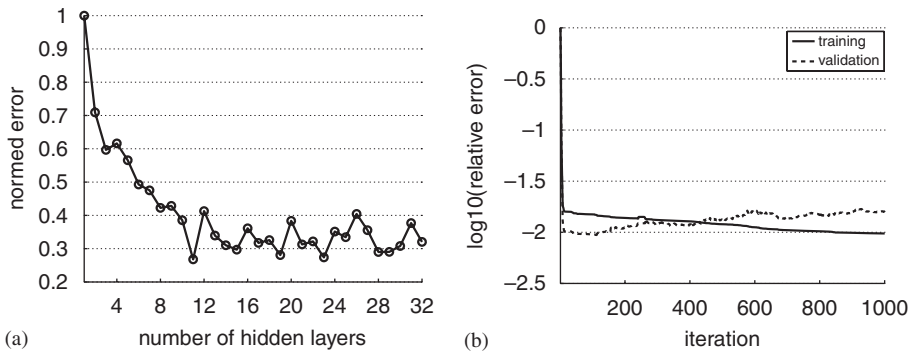

Figure 5. Variation of the trained MLP error with the number of units in the hidden layer (a). Error histories on the training and validation sets during the training of the gradient-enhanced MLP (b).

The rise in efficiency is modest (about 0.6%) and is only limited by the low-dimensionality of the design space.

The second exercise is a realistic multi-objective optimization in a high-dimensional design space. The task was to maximize the stage isentropic efficiency of the NASA Rotor 37 at the working conditions, with a minimum variation of the stage pressure ratio around the datum value of the base rotor geometry. The design space was defined by 15 parameters; there were 5 types

DOI: 10.1002/fld

Table I. History of relative errors (percent) between the optima based on the gradient-enhanced MLP approximations and the direct simulated values.

	Errors		
Iterations	SA	GA1	GA2
1	4.12	4.47	0.83
2	2.43	0.45	0.45
3	0.20	0.25	1.44
4	2.40	0.94	0.20
5	1.43	0.86	0.54
6	0.46	0.54	0.43

Values are given for all three optimizers used.

of design perturbations (sweep, lean, skew, leading and trailing edge re-cambering) applied at 3 different locations along the blade span (40%, 80% of height and tip). The spanwise variation of the design perturbations is interpolated from the set locations using cubic splines; the effect on the hub re-design is small but not zero. The initial DoE consisted on 300 LP_τ points, of which 271 were kept; the rest were rejected because of failed computations.

The number of hidden layer units necessary to model the DoE was assessed by training the MLP for a range of sizes of the hidden layer, Figure 5(a). The optimal number was found to be 10; including more hidden layer units into the MLP leads to no noticeable improvement in accuracy. The number of iterations necessary to train the MLP was determined by using about a quarter of the DoE for validation and allowing the training to proceed for a large number of iterations. Figure 5(b) shows that after a number of iterations the error on the training set is reduced at a slow rate while the error on the validation set begins to increase. At this point the training is stopped.

The optimization procedure was iterative and based on the off-line trained MLP. At each iteration, the MLP was trained and then independently used as a RSM by three global-search optimizers; SA and two different GAs (code-named GA1 and GA2), all three available from Soft [11]. At each iteration, the Pareto fronts of non-dominated points resulting from the three independent optimizations were combined and used to update the response/gradients DoE with simulated flows. The errors made at each step between the *best* optima (characterised by the highest efficiency) from the Pareto fronts of MLP predictions and their *true* values from flow simulations were recorded in Table I. After six iterations, the errors made by all three optimizers dropped to within 0.5% and the convergence of optimization was considered achieved. Because the optimization update was achieved using Pareto fronts and not single-objective values, the optimization convergence was followed in terms of the Pareto front stabilization rather than the convergence of a single update, hence the error adopted to monitor convergence. At convergence, the DoE contained 346 points.

This optimization produced a solution with an rise in isentropic efficiency at the working condition of 2.55% and a variation of pressure ratio of 1.05%. Computing the entire characteristic of this re-design, the rise in isentropic efficiency was found to be consistent across all the back-pressure conditions, from choke to stall, Figure 6(a). Also, the deviation of the pressure ratio from the datum was shown to be small over the entire characteristic, Figure 6(b). This optimal geometry is referred to as the *optim 2* solution.

DOI: 10.1002/fld

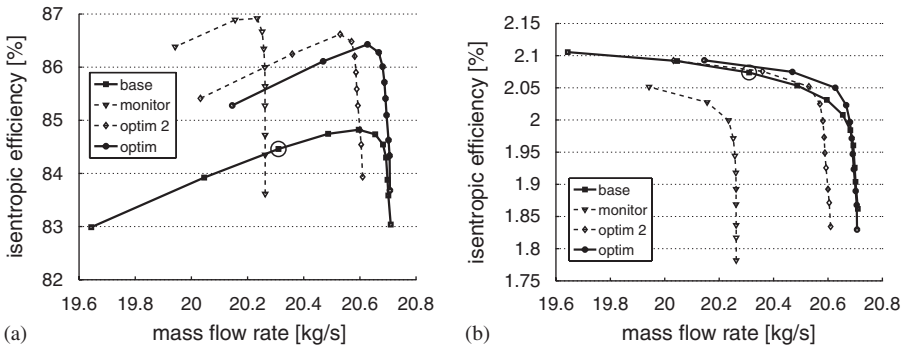

Figure 6. The isentropic efficiency characteristics of the datum Nasa Rotor 37 geometry, the *monitor* local optimum and the two global optimal re-designs *optim* and *optim 2* (a). The pressure ratio characteristics of the same configurations (b). The *monitor* optimum gives the highest raise in efficiency but has unacceptable variations in the pressure ratio and choke massflow. The *optim 2* solution has a lower raise in efficiency, but an acceptable variation of pressure ratio. The *optim* solution has a slightly lower efficiency raise than *optim 2*, but is the acceptable re-design. The optimized working point is marked with a circle on the characteristic.

The *optim 2* solution is not valid from an engineering point of view because the choke massflow of the re-design deviates by 0.5% from the datum, which is unacceptably high. To correct this, a second optimization was performed, adding an extra objective to minimize, defined as the deviation of the choke massflow from the base geometry figure. The design space and the optimization procedure were similar to those of the *optim 2* optimization. After convergence of the optimization procedure, the computed characteristic confirmed the validity of the found optimum, referred to as the *optim* solution. The rise in efficiency is 2.33%, slightly smaller than the *optim 2* value but likewise consistent across the characteristic, Figure 6(a). The deviation of stage pressure ration was 1.10% at the working condition and within the same range at all conditions, Figure 6(b). The main difference from the previous result is the small deviation (0.015%) of the choke massflow from the baseline.

To compare results with an established method, a separate single-objective local gradient-based optimization of efficiency was performed at working conditions in the same design space using a Sequential Quadratic Programming search [11]. This algorithm uses the objective and gradients output by the flow simulations directly, without recourse to a RSM. The search was started at the baseline design and had to be unconstrained in order to achieve a local solution significantly different from the datum. The result, referred to as *monitor*, was much better in terms of the raise in efficiency (3%) than *optim* or *optim 2*, but the engineering acceptability is violated by deviations in pressure ratio and choke massflow of 3 and 3.5%, respectively. The characteristics of this re-design are also shown in Figure 6(a) and (b).

Figure 7 depicts three-dimensional Pareto front of the second *optim* optimization. The non-dominated solutions in the space defined by isentropic efficiency, pressure ratio variation and choke massflow variation are shown in projection on the efficiency/pressure ratio variation plane, Figure 7. From the Pareto front, several candidates to the final optimal design were selected and the decision for the *optim* solution was made under the assumption that 1% is an acceptable level for the pressure ratio deviation from baseline. For comparison, Figure 7 also indicates the position of the *optim 2* solution.

DOI: 10.1002/fld

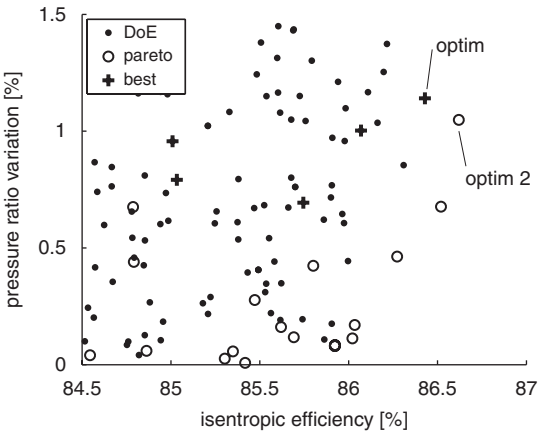

Figure 7. The 3D Pareto front of the global optimization leading to the *optim* solution projected on the plane defined by the isentropic efficiency and the variation in pressure ratio. Shown are points in the DoE (•), members of the front (○) and the candidates to the optimal solution (+). The *optim* and *optim 2* solutions are also indicated.

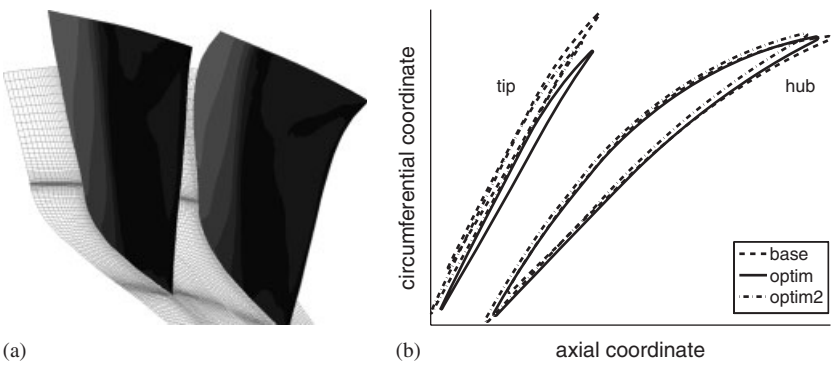

Figure 8. Depicted in (a) are the base (left) and the *optim* (right) geometries, with the shading representing the variation of the static pressure fields on the suction sides. The shock and flow separation achieved through re-design are responsible for the higher isentropic efficiency. The main effects of the *optim* and *optim 2* re-designs are a forward blade sweep, shown through the change in the blade profile at hub and tip (b).

Finally, Figure 8(a) depicts the geometries of the datum blade and the *optim* global solution. The main geometric effect of the re-design is a forward sweep at the tip, which accounts for the rise in efficiency by a delayed shock (and resulting flow separation) in the optimal design; this is in line with the results in [17]. Also, Figure 8(b) depicts the profiles of the blade at hub and tip, showing similar effects of redesign in both the *optim* and the *optim 2* solutions.

DOI: 10.1002/fld

6. CONCLUSION

The paper re-formulates a gradient-enhanced multi-layer perceptron (MLP) and presents a realistic multi-objective shape optimization of turbomachinery blading using the perceptron as a response surface approximations.

The formulation is carried out using algebraic notations; this approach has the advantage that is easy to follow and leads naturally to a fast computer code implementation using high-performance BLAS for the matrix and vector operations. The paper explains how the forward prediction and back-propagation schemes of the perceptron are modified to include gradient information. The correctness and robustness of the gradient-enhanced perceptron are assessed in a suite of simple tests.

The enhanced MLP is successfully used in the multi-objective optimization of the transonic Nasa Rotor 37. The stage isentropic efficiency is raised significantly all across the full-speed characteristic, while the pressure ratio variation about the baseline figure is minimal. An interesting case from the engineering point of view is made by the addition of the deviation of choke massflow from the datum figure as an extra objective to minimize. The found optimum is characterized by geometric features in line with the results from other researchers.

ACKNOWLEDGEMENTS

This research was partially supported by an MCDO project funded by the U.K. Department for Trade and Industry and Rolls-Royce plc, and coordinated by S. Shahpar, Y. Ho and L. Lapworth.

REFERENCES

1. Myers RH, Montgomery DC. *Response Surface Methodology*: *Process and Product Optimization Using Designed Experiments*. Wiley Series in Probability and Statistics. Wiley: New York, 1995.
2. Jones DR. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization* 2001; **21**:345–383.
3. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* 1998; **13**:455–492.
4. Leary SJ, Bhaskar A, Keane AJ. A derivative based surrogate model for approximating and optimizing the output of an expensive computer simulation. *Journal of Global Optimization* 2004; **30**:39–58.
5. Chung HS, Alonso JJ. Using gradients to construct Cokriging approximation models for high-dimensional design optimization problems. *AIAA Paper 2002-0317*, 2002.
6. Giannakoglou KC, Papadimitriou D, Kampolis IC. Aerodynamic shape design using evolutionary algorithms and new gradient-assisted metamodels. *Computer Methods in Applied Mechanics and Engineering* 2006; **195**: 6312–6329.
7. Demeulenaere A, Hirsch C. Application of multipoint and multistage optimization to the design of turbomachinery blades. *Proceedings of the 6th European Conference on Turbomachinery*, Lille, France, 7–11 March 2005.
8. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press: Oxford, 1995.
9. Muselli M. A theoretical approach to restart in global optimization. *Journal of Global Optimization* 1997; **10**:1–16.
10. Sobol IM. On the systematic search in a hypercube. *SIAM Journal on Numerical Analysis* 1979; **16**(5):790–793.
11. Shahpar S. SOFT: a new design and optimisation tool for turbomachinery, evolutionary methods for design, optimisation and control. *Proceedings of the Evolutionary Methods for Design*, *Optimisation and Control*, *CIMNE*, Barcelona, Spain, 2002.
12. Dunham J (ed.). CFD validation for propulsion system components. *AGARD Advisory Report 355*. ISBN: 92-836-1075-X.
13. Shahpar S, Lapworth BL. PADRAM: parametric design and rapid meshing system for turbomachinery optimisation. *ASME Paper GT-2003-38398*, 2003.

DOI: 10.1002/fld

14. Lapworth BL, Shahpar S. Design of gas turbine engines using CFD. *Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering*, *ECCOMAS*, Jyväskylä, Finland, 2004.
15. Giles MB, Duta MC, Müller J-D, Pierce NA. Algorithm developments for discrete adjoint methods. *AIAA Journal* 2003; **41**(2):198–205.
16. Duta MC, Giles MB. The use of automatic differentiation for adjoint CFD codes. *Proceedings of the European Conference on Computational Fluid Dynamics*, Egmond aan Zee, The Netherlands, 5–8 September 2006.
17. Hah C, Puterbaugh SL, Wadia AR. Shock structure and secondary flow field inside transonic compressor rotor through aerodynamic sweep. *ASME Paper 99-GT-561*, 1999.

DOI: 10.1002/fld