Available online at www.sciencedirect.com

ScienceDirect

Applied Mathematical Modelling 33 (2009) 1573–1583

APPLIED MATHEMATICAL MODELLING

www.elsevier.com/locate/apm

On the selection of the most adequate radial basis function

Humberto Rocha [*]

Portuguese Catholic University, Pólo de Viseu, 3504-505 Viseu, Portugal

Received 13 September 2007; received in revised form 7 January 2008; accepted 13 February 2008
Available online 10 March 2008

Abstract

Radial basis function (RBF) methods can provide excellent interpolants for a large number of poorly distributed data points. For any finite data set in any Euclidean space, one can construct an interpolation of the data by using RBFs. However, RBF interpolant trends between and beyond the data points depend on the RBF used and may exhibit undesirable trends using some RBFs while the trends may be desirable using other RBFs. The fact that a certain RBF is commonly used for the class of problems at hand, previous good behavior in that (or other) class of problems, and bibliography, are just some of the many valid reasons given to justify a priori selection of RBF. Even assuming that the justified choice of the RBF is most likely the correct choice, one should nonetheless confirm numerically that, in fact, the most adequate RBF for the problem at hand is the RBF chosen a priori. The main goal of this paper is to alert the analyst as to the danger of a priori selection of RBF and to present a strategy to numerically choose the most adequate RBF that better captures the trends of the given data set. The wing weight data fitting problem is used to illustrate the benefits of an adequate choice of RBF for each given data set.

Keywords: Radial basis function; Data fitting; Cross-validation; Principal component regression

1. Introduction

Radial basis function (RBF) methods can provide excellent interpolants for high dimensional data sets of poorly distributed data points (scarce and unevenly distributed points). For any finite data set in any Euclidean space, one can construct an interpolation of the data by using RBFs, even if the data points are unevenly and sporadically distributed in a high dimensional Euclidean space. There is a wide range of applications where RBF interpolation methods can be successfully applied, such as aeronautics, meteorology and medical imaging (see [1–5]). However, RBF interpolant trends between and beyond the data points depend on the RBF used and may exhibit undesirable trends using some RBFs while the trends may be desirable using other RBFs.

The fact that a certain RBF is commonly used for the class of problems at hand, previous good behavior in that (or other) class of problems, and bibliography, are just some of the many valid reasons given to justify a

[*] Tel.: +351 239109410.
E-mail address: hrocha@mat.uc.pt

0307-904X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.apm.2008.02.008

priori selection of RBF. Even assuming that the justified choice of the RBF is most likely the correct choice, i.e., that the RBF picked is the most adequate and the one that produces the "best" model, one should nonetheless confirm numerically that, in fact, the most adequate RBF for the problem at hand is the RBF chosen a priori.

The main goal of this paper is to alert the analyst as to the danger of a priori choice of RBF and to present a strategy to numerically choose the most adequate RBF that better captures the trends of the given data set. The wing weight data fitting problem is used to illustrate the benefits of an adequate choice of RBF for each given data set.

The paper is organized as follows. Section 2 gives a brief description of RBF interpolation problems. Section 3 introduces cross-validation for model parameter tuning. In Section 4, wing weight data fitting is used to illustrate the benefits of an adequate choice of RBF for each given data set. Section 5 includes the concluding remarks.

2. RBF interpolation problems

Let $f(\mathbf{x})$ be a real-valued function of the input vector $\mathbf{x}$ defined on a subset Ω of $\mathbb{R}^n$ such that the value of f at N input vectors $\mathbf{x}^j, j = 1, \ldots, N, f(\mathbf{x}^j)$, is given. The goal is to construct an estimation model $g(\mathbf{x})$ such that $g(\mathbf{x}^j) = f(\mathbf{x}^j)$ for $j = 1, \ldots, N$. The interpolation requirement can be satisfied by RBF interpolation.

Interpolation functions generated from a RBF $\varphi(t)$ can be represented in the following form:

$$g(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j \varphi(\|\mathbf{x} - \mathbf{x}^j\|), \tag{1}$$

where $||\mathbf{x} - \mathbf{x}^j||$ denotes the parameterized distance between $\mathbf{x}$ and $\mathbf{x}^j$ defined as

$$||\mathbf{x} - \mathbf{x}^j|| = \sqrt{\sum_{i=1}^{n} |\theta_i|(x_i - x_i^j)^2},$$

and $\theta_1, \ldots, \theta_n$ are scalars (see [1]).

The most popular examples of RBF [6–8] are cubic spline $\varphi(t) = t^3$, thin plate spline $\varphi(t) = t^2 \ln t$, multiquadric $\varphi(t) = \sqrt{1 + t^2}$, and Gaussian $\varphi(t) = \exp(-t^2)$ (see Fig. 1). These RBFs can be used to model cubic, almost quadratic, and linear growth rates, as well as exponential decay, of the response for trend predictions.

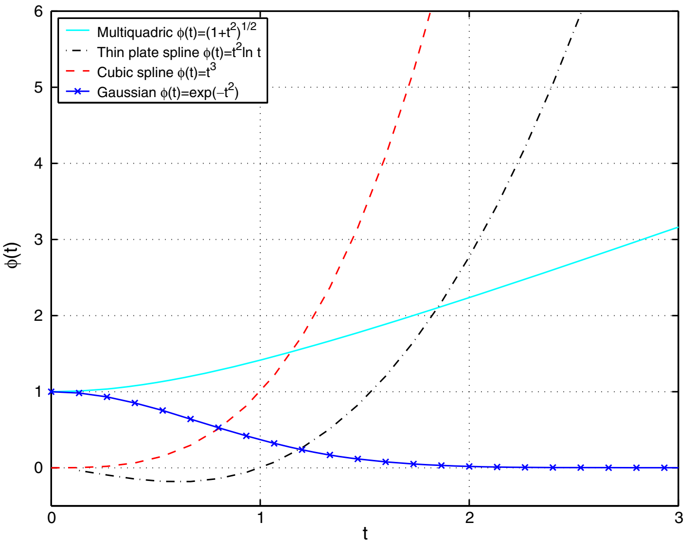

Fig. 1. Graphs of commonly used radial basis functions.

For fixed parameters θ_i, the coefficients $\alpha_1, \ldots, \alpha_N$ in Eq. (1) can be calculated by solving the following linear system of interpolation equations:

$$\sum_{j=1}^{N} \alpha_j \varphi(||\mathbf{x}^k - \mathbf{x}^j||) = f(\mathbf{x}^k), \quad \text{for } k = 1, \ldots, N. \tag{2}$$

One can rewrite Eq. (2) in matrix form as

$$\Phi \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}^1) \\ f(\mathbf{x}^2) \\ \vdots \\ f(\mathbf{x}^N) \end{pmatrix}, \tag{3}$$

where Φ is the interpolation matrix defined as

$$\Phi = \begin{pmatrix} \varphi(||\mathbf{x}^1 - \mathbf{x}^1||) & \varphi(||\mathbf{x}^1 - \mathbf{x}^2||) & \ldots & \varphi(||\mathbf{x}^1 - \mathbf{x}^N||) \\ \varphi(||\mathbf{x}^2 - \mathbf{x}^1||) & \varphi(||\mathbf{x}^2 - \mathbf{x}^2||) & \ldots & \varphi(||\mathbf{x}^2 - \mathbf{x}^N||) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(||\mathbf{x}^N - \mathbf{x}^1||) & \varphi(||\mathbf{x}^N - \mathbf{x}^2||) & \ldots & \varphi(||\mathbf{x}^N - \mathbf{x}^N||) \end{pmatrix}.$$

A unique interpolant is guaranteed for multiquadric and Gaussian RBFs (i.e., Φ is a nonsingular matrix) even if the input vectors $\mathbf{x}^j$ are few and poorly distributed, provided only that the input vectors are all different when $N > 1$. However, for cubic and thin plate spline RBFs, Φ might be singular. See [6] for an example of singular Φ when $\varphi(t) = t^3$. If $\varphi(t) = t^2 \ln t$, we can easily find an example where the interpolation matrix Φ is singular for a nontrivial set of distinct points $\mathbf{x}^1, \ldots, \mathbf{x}^N$. For example, let $\mathbf{x}^2, \ldots, \mathbf{x}^N$ be any distinct points on the sphere centered at $\mathbf{x}^1$ with radius 1. For this set of points, the first row and column of Φ consist of zeros, which implies the singularity of Φ.

An easy way to avoid this problem on the cubic and thin plate spline RBF interpolants is to add low-degree polynomials to interpolation functions in (1) and formulate an interpolation problem with constraints. That is, let $p(\mathbf{x}) = \sum_{j=1}^{M} \beta_j p_j(\mathbf{x})$, where $p_1, \ldots, p_M$ form a basis of algebraic polynomials in $\mathbb{R}^n$ with degree at most m. Then interpolation functions are of the following form:

$$g(\mathbf{x}) = p(\mathbf{x}) + \sum_{j=1}^{N} \alpha_j \varphi(||\mathbf{x} - \mathbf{x}^j||). \tag{4}$$

The M extra degrees of freedom in $g(\mathbf{x})$ can be eliminated by forcing the following M constraints:

$$\sum_{j=1}^{N} \alpha_j p_k(\mathbf{x}^j) = 0, \quad \text{for } k = 1, \ldots, M, \tag{5}$$

which has the following matrix form:

$$\mathbf{P} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = 0,$$

where

$$\mathbf{P} = \begin{pmatrix} p_1(\mathbf{x}^1) & \ldots & p_1(\mathbf{x}^N) \\ \vdots & \ddots & \vdots \\ p_M(\mathbf{x}^1) & \ldots & p_M(\mathbf{x}^N) \end{pmatrix}.$$

The interpolation equations using $g(\mathbf{x})$ in Eq. (4) become

$$\sum_{j=1}^{N} \alpha_j \varphi(\|\mathbf{x}^k - \mathbf{x}^j\|) + \sum_{j=1}^{M} \beta_j p_j(\mathbf{x}^k) = f(\mathbf{x}^k), \quad \text{for } k = 1, \ldots, N. \tag{6}$$

Combining Eqs. (5) and (6), we obtain the following equation for the constrained RBF interpolation in matrix form:

$$\begin{pmatrix} \Phi & \mathbf{P}^{\mathrm{T}} \\ \mathbf{P} & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \\ \beta_1 \\ \vdots \\ \beta_M \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_N \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{7}$$

The key results on solvability of RBF interpolations related to the four RBFs shown in Fig. 1 are the following:

(1) As seen before, Eq. (3) is always solvable if $\varphi(t) = \sqrt{1+t^2}$ or $\varphi(t) = \exp(-t^2)$.
(2) For quadratic polynomials ($m = 2$), Eq. (7) is solvable if $\varphi(t) = t^3$ or $\varphi(t) = t^2 \ln t$, provided that the input vectors $\mathbf{x}^1, \ldots, \mathbf{x}^N$ do not fall into the zero set of the polynomial.

The proofs of second statements are based on mathematical concepts called (conditionally) positive definiteness and their Schoenberg–Micchelli characterizations [9,10] as described by Schaback and Wendland [11].

The constructed interpolant $g(\mathbf{x})$ in Eq. (1) depends on "subjective" choice of $\varphi(t)$, and model parameters $\theta_1, \ldots, \theta_n$. While one can try all the possible choices of $\varphi(t)$ in search of a desirable interpolant, there are infinitely many choices for $\theta_1, \ldots, \theta_n$. One could, however, use cross-validation to choose an optimal value of $\theta_1, \ldots, \theta_n$ that yield an interpolant $g(\mathbf{x})$ with the most accurate trend prediction.

3. Model parameter tuning

RBF interpolation models use the parameterized distance:

$$\|\mathbf{x} - \mathbf{x}^j\| = \sqrt{\sum_{i=1}^{n} |\theta_i| (x_i - x_i^j)^2},$$

where $\theta_1, \ldots, \theta_n$ are scalars. Mathematically, one could pick any fixed set of $\theta_1, \ldots, \theta_n$ and construct the interpolation function for the given data. However, two different sets of $\theta_1, \ldots, \theta_n$ will lead to two interpolation models that behave very differently between the input vectors $\mathbf{x}^1, \ldots, \mathbf{x}^N$. Model parameter tuning for RBF interpolation aims at finding a set of parameters $\theta_1, \ldots, \theta_n$ that results in the best prediction model of the unknown response based on the available data. Cross-validation (CV) [12–17] and maximum likelihood estimation [18–20] are two statistical methods for tuning the model parameters $\theta_1, \ldots, \theta_n$ for best prediction models.

CV can be used for general model parameter tuning, while maximum likelihood estimation can only be applied for density function parameter estimation. Cross-validation is introduced in the next section.

3.1. Model parameter tuning by CV

Other metrics instead of fitting errors must be used to determine which basis function $\varphi(t)$ and what scaling parameters θ_i are most appropriate to model the response function $f(\mathbf{x})$, because RBF interpolation method yields a fitting function $g(\mathbf{x})$ whose value at $\mathbf{x}^k$ is exactly $f(\mathbf{x}^k)$ for $k = 1, \ldots, N$. One can always try to obtain values of $f(\mathbf{x})$ at some additional data points $\mathbf{x}^{N+1}, \ldots, \mathbf{x}^{\overline{N}}$ and use the prediction errors $|g(\mathbf{x}^k) - f(\mathbf{x}^k)|$ for $k = N+1, \ldots, \overline{N}$ to assess the prediction accuracy of $g(\mathbf{x})$, but this technique is often impractical and always expensive. The prediction accuracy can be used as a criterion for choosing the best basis function $\varphi(t)$ and parameters θ_i.

Without additional sample points, CV [14,16] was proposed to find $\varphi(t)$ and θ_i that lead to an approximate response model $g(\mathbf{x})$ with optimal prediction capability and proved to be effective [15,17]. The leave-one-out CV procedure is usually used in model parameter tuning for RBF interpolation (see [15,21]).

3.1.1. Leave-one-out cross-validation for RBF interpolation

- Fix a set of parameters $\theta_1, \ldots, \theta_n$.
- For $j = 1, \ldots, N$, construct the RBF interpolant $g_{-j}(\mathbf{x})$ of the data points $(\mathbf{x}^k, f(\mathbf{x}^k))$ for $1 \leqslant k \leqslant N, k \neq j$.
- Use the following CV root mean square error as the prediction error:

$$E^{\mathrm{CV}}(\theta_1, \ldots, \theta_n) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(g_{-j}(\mathbf{x}^j) - f(\mathbf{x}^j))^2}. \tag{8}$$

Remark. One could also use other forms of CV errors such as the CV average absolute error: $\frac{1}{N}\sum_{j=1}^{N}|g_{-j}(\mathbf{x}^j) - f(\mathbf{x}^j)|$.

In the case that each $\mathbf{x}^j$ has a close neighbor $\mathbf{x}^k$ $(k \neq j)$ in the space of input variables and $f(\mathbf{x})$ is a smooth function, the cross-validation error (E^{CV}) is not a meaningful measure of the prediction accuracy of the fitting model because $g_{-j}(\mathbf{x}^k) = f(\mathbf{x}^k)$ implies $g_{-j}(\mathbf{x}^j) \approx f(\mathbf{x}^j)$ due to small values of $||\mathbf{x}^k - \mathbf{x}^j||$ and $|f(\mathbf{x}^k) - f(\mathbf{x}^j)|$. Therefore, if the leave-one-out CV error is used as a criterion for model parameter tuning, then analysts must be warned when undesirable clustering of $\mathbf{x}^j$ occurs.

The goal of model parameter tuning by CV is to find $\theta_1, \ldots, \theta_n$ that minimize the CV error $E^{\mathrm{CV}}(\theta_1, \ldots, \theta_n)$ so that the interpolation model has the highest prediction accuracy when CV error is the measure. It is worth pointing out that, typically, it is difficult to minimize $E^{\mathrm{CV}}(\theta_1, \ldots, \theta_n)$ numerically because $E^{\mathrm{CV}}(\theta_1, \ldots, \theta_n)$ is a highly nonlinear and nonconvex function. One could make the model parameter tuning much easier by assuming $\theta_1 = \ldots = \theta_n$, which reduces the problem to unconstrained minimization of a univariate function (see [15]). This approach has the obvious benefit of dealing with a unidimensional optimization problem but the disadvantage of not using all different θ_i. Different θ_i allow the model parameter tuning to scale each variable x_i based on its significance in modeling the variance in the response, thus, have the benefit of implicit variable screening built in the model parameter tuning.

4. Wing weight data fitting problem

The wing weight fitting problem is used here to demonstrate the need for an adequate selection of the RBF to use in the interpolation model. This problem was used to show feasibility of using approximation methods in construction of a wing weight estimation formula for conceptual design of subsonic transports [1–3]. For system analysis of conceptual design of aircraft, one important task is to resize a conceptual aircraft for a mission analysis. To conduct a mission analysis of a resized aircraft, system analysts must estimate the gross takeoff weight w_{to} of the aircraft. Specifically, one commonly resized component of aircraft is wing. As a result, system analysts need a relationship between the wing weight w and sizing parameters of wing (such as s (plan area of the wing), b (span), λ (taper ratio), and Λ (sweep angle)).

The wing weight data set was collected over a period of time and will be expanded when new weight statements of subsonic transports become available. System analysts at NASA Ames Research Center were able to collect weight information of 41 subsonic transports including Boeing 747, Douglas DC-7C, Fokker F-28 twin engine jet liner, and Lockheed C-130B cargo aircraft (see Fig. 2). Such dynamic characteristics of the data set requires a relatively easy way to generate a wing weight estimation model to capture the trend in the updated data set, when currently used regression models, such as the models suggested by Ardema et al. [22], become inadequate as estimation models.

Fig. 2. Examples of subsonic transports in the data set.

Each wing weight data point consists of the actual wing weight w and relevant key configuration parameters: A, b, c_m, c_r, c_t, s, $[t/c]_m$, t_r, t_r/c_r, t_t, t_t/c_t, w_{to}, λ, Λ, and μ. These parameters can be regrouped in three categories: (1) wing geometry parameters including chord length at root (c_r), chord length at tip (c_t), span (b), reference area (s), thickness at root (t_r), thickness at tip (t_t), and sweep angle ($\Lambda = |90 - \Lambda_0|$); (2) wing aerodynamics parameters including taper ratio ($\lambda = c_t/c_r$), aspect ratio ($A = b^2/s$), mean chord of wing ($c_m = s/b$), thickness-to-chord ratio at root (t_r/c_r), and thickness-to-chord ratio at tip (t_t/c_t); and (3) wing structure parameters including gross takeoff weight of aircraft (w_{to}) and wing load factor (μ). Fig. 3 shows the wing geometry parameters for a trapezoidal approximation of the actual wing. A detailed explanation of wing configuration parameters can be found in Raymer's book [23].

The goal is to construct a weight estimation model $\bar{w} \approx w$, where $\bar{w}$ is a function of all or a subset of the configuration parameters. Notice that some configuration parameters are related, e.g., $t_t/c_t + t_r/c_r = 2[t/c]_m$ and $A = b^2/s$. There are many different ways to select a set of independent configuration parameters, leading to a completely different regression model if a general approximation model is used. Similarly, each choice of configuration parameters as input variables of a general approximation model (such as radial basis function models) will lead to a new regression model. As a consequence, one must exercise caution in determination of which configuration parameters should be used as the input variables and which RBF should be used.

Because we use interpolation methods (RBF interpolation), each wing weight estimation model will reproduce the actual wing weight for 41 subsonic transports. However, such an exact fit of the data has no useful purpose for the sizing of wing in conceptual design phase. System analysts are mostly interested in whether an estimation model captures the weight growth trends "correctly" between and beyond the known data points. There is no physics-based criterion for verification of a correct solution; instead, expert opinions determine whether a mathematical solution is useful in practice. For wing weight estimation, a desirable approximation should have the following properties: w is an increasing function with respect to each of b, s, λ, and Λ; and w is

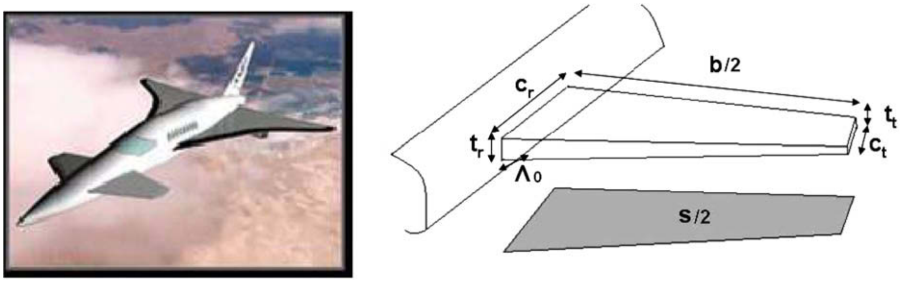

Fig. 3. Airplane wing geometry parameters.

a decreasing function with respect to t_r/c_r. These properties are derived from simple engineering rules on the relationships between the wing weight and each of the five key configuration parameters.

For multivariate data fitting problems, it is not easy to decide what should be potential input variables. For the wing weight approximation problem, the input variables are usually the sizing parameters required in conceptual design of aircraft. We used two sets of variables (see Refs. [1,3]): a set containing all the configuration parameters of wing (including the ratios) and a set of the variables used by engineers in the construction of their best empirical model [4], i.e., $\mathbf{x}$ is a vector of $n = 8$ configuration parameters: b, s, t_r/c_r, λ, Λ, μ, w_{to}, and t_t/c_t. The goal of the first set was to justify the engineer's choice of their current set of parameters, while the second set was used to construct a better approximation for the problem at hand. One will focus on the second goal of improving the existing approximation and will only consider here the second set of $n = 8$ configuration parameters.

The first step of the approximation procedure is to find out which of the variables are important for a wing weight estimation model and whether there is any collinearity of the input variables. Principal component analysis (PCA) can be used successfully for that purpose [24, Section 3.6]. By applying PCA to the input vectors $\mathbf{x}^1, \ldots, \mathbf{x}^N$, we can treat the response as a function defined on a feature space $\mathbb{R}^r$ with a reduced dimension $r\ (\leqslant n)$, solve the approximation problem by fitting the transformed data in the feature space, and then recover the approximate response in the original input space (see [24, Section 3.6]).

One could also use PCA to transform the data into a feature space and use an interpolation method in the feature space. The principal component regression using RBF interpolation [1–3] is such an example. Specifically, let $\mathbf{v}$ be the vector with r components $(\hat{\mathbf{x}} - \mathrm{ave}(\hat{\mathbf{x}}))^{\mathrm{T}}\mathbf{u}_1, \ldots, (\hat{\mathbf{x}} - \mathrm{ave}(\hat{\mathbf{x}}))^{\mathrm{T}}\mathbf{u}_r$. Then the N data points in the $\mathbf{x}$-space can be mapped to N points in the $\mathbf{v}$-space. The corresponding problem of fitting N data points $(\mathbf{v}^1, y_1), \ldots, (\mathbf{v}^N, y_N)$ can be solved by using a standard interpolation method. If the solution of the data fitting problem in the $\mathbf{v}$-space is $\hat{g}(\mathbf{v})$, then the corresponding fitting of the data points in the $\mathbf{x}$-space is

$$\hat{f}(\mathbf{x}) = \hat{g}((\hat{\mathbf{x}} - \mathrm{ave}(\hat{\mathbf{x}}))^{\mathrm{T}}\mathbf{u}^1, \ldots, (\hat{\mathbf{x}} - \mathrm{ave}(\hat{\mathbf{x}}))^{\mathrm{T}}\mathbf{u}^r), \tag{9}$$

where $\hat{\mathbf{x}}$ is the scaled version of $\mathbf{x}$ and its ith component is $\hat{x}_i = x_i/\sigma_i$ with σ_i an estimation of its standard deviation calculated from the data, $\mathrm{ave}(\hat{x}_i) = \frac{1}{N}\sum_{k=1}^{N}\hat{x}_i^k$ is the ith component of $\mathrm{ave}(\hat{\mathbf{x}})$, and $\mathbf{u}^1, \ldots, \mathbf{u}^n$ are the unit eigenvectors corresponding to the spectral decomposition of the covariance matrix $\mathbf{C}$ of the scaled input vectors $\hat{\mathbf{x}}^1, \ldots, \hat{\mathbf{x}}^N$:

$$\mathbf{C} = \frac{1}{N-1}\sum_{j=1}^{N}[\hat{\mathbf{x}}^j - \mathrm{ave}(\hat{\mathbf{x}})][\hat{\mathbf{x}}^j - \mathrm{ave}(\hat{\mathbf{x}})]^{\mathrm{T}}.$$

Note that different choices of r generate different interpolants $\hat{f}(\mathbf{x})$ even if the same RBF $\varphi(t)$ is used. The option of choosing r allows one to explore the degree of freedom in the input space that is most appropriate to describe the relationship between the input vector and the response. The restriction on r is that $\mathbf{v}^1, \ldots, \mathbf{v}^N$ must be mutually distinct. In practice, it is not easy to choose r. The key issue is whether one wants to extract a functional relationship between the change of the scaled input vector $\hat{\mathbf{x}}$ along a direction $\mathbf{u}^j$ and the response even if the existing input vectors $\hat{\mathbf{x}}^1, \ldots, \hat{\mathbf{x}}^N$ have a small variation along the direction $\mathbf{u}^j$. One can try all possible choices for r and use a systematic evaluation process to down select the approximation that is most appropriate for the analysis task at hand. In order to do that we suggest the following automatic principal component regression (PCR) procedure.

4.1. Automatic PCR procedure

- Compute the eigenvalues of the covariance matrix $\mathbf{C}$ and let $\bar{r}$ be the number of positive eigenvalues.
- Choose an integer $r_{\min} < \bar{r}$ such that there is still a functional relationship between the transformed input vectors in the $r_{\min}$-dimensional feature space and the response. For each integer r from $r_{\min}$ to $\bar{r}$, minimize the CV error for the given interpolation model in the r-dimensional feature space.
- Choose r corresponding to the smallest value of the minimized CV errors and select the corresponding interpolant as $\hat{g}(\mathbf{v})$.

- Reconstruct the wing weight approximation in the **x**-space by using Eq. (9).

We use MATLAB code **fminsearch**, an implementation of the Nelder–Mead [25] multidimensional search algorithm, to minimize the CV error $E^{\text{CV}}(\theta_1, \ldots, \theta_n)$ in Eq. (8) and to find the best model parameters $\theta_1, \ldots, \theta_n$. The local optimal solution generated by MATLAB code **fminsearch** for minimization of the CV error is very reliable but also very sensitive to the initial guess. Multiple initial guesses were used for searching a global minimizer of the CV error by **fminsearch**.

The CV error of an interpolation model can be a useful and objective tool to help analysts decide which model is better. In Table 1, we have the CV errors of various interpolation models (RBF interpolation models with different basis functions). In particular, the first row of Table 1 has the CV errors for interpolation models in the **x**-space, while the other rows show the CV errors in the r-dimensional feature spaces. The above automatic PCR procedure can help to reduce the PCR interpolants to one for each interpolation model. In the case of multiquadric RBF interpolation with $n = 8$, $\bar{r} = n$, and $r_{\min} = 4$, the best PCR interpolant is generated by using $r = 7$. This PCR interpolant is indeed the most desirable wing weight prediction model for subsonic transports and proved to have the overall best trends for wing weight prediction by inspecting the five types of two dimensional plots for all 41 baseline configurations [1–3].

The CV error of a constructed approximation can be useful to help analysts decide which approximation is better. For wing weight data fitting with a fixed type of interpolation models, the "best" approximation model among all the approximations generated by PCR (for a range of r) usually corresponds to the smallest value of minimized CV errors. Therefore, it makes sense to use the automatic PCR procedure for choosing the best PCR for any fixed interpolation model.

Because the difference between the smallest and the largest wing weight values is considerable, one must verify whether the CV error value E^{CV} is dominated by the prediction errors at $\mathbf{x}^j$ with large wing weight values. In other words, we should compute the relative CV prediction errors for each data point, defined as $(g_{-j}(\mathbf{x}^j) - f(\mathbf{x}^j))/g_{-j}(\mathbf{x}^j)$ in Table 2, and check if the CV optimization attempts to minimize the prediction errors $(g_{-j}(\mathbf{x}^j) - f(\mathbf{x}^j))^2$ for large $f(\mathbf{x}^j)$. One indicator for such a biased minimization of E^{CV} is that the relative errors for large $g_{-j}(\mathbf{x}^j)$ (or $f(\mathbf{x}^j)$) are smaller than those for small $g_{-j}(\mathbf{x}^j)$. But we do not see such a biased minimization of E^{CV} in Table 2, which shows the relative errors $(g_{-j}(\mathbf{x}^j) - f(\mathbf{x}^j))/g_{-j}(\mathbf{x}^j)$ for $j = 1, \ldots, N$ when CV is applied to multiquadric PCR with $r = 7$. The CV prediction errors can also be used as a tool to help on the decision of the best approximation.

Fig. 4 shows the leave-one-out CV predictions $g_{-j}(\mathbf{x}^j)$ $(j = 1, \ldots, N)$ for the multiquadric PCR with $r = 7$ and the best engineer model (geometry model). It is clear that the multiquadric PCR with $r = 7$ has much more accurate predictions than the geometry model.

Therefore, this suggests a numerical decision tool for the choice of the most adequate RBF to be used.

4.2. Numerical choice of the most adequate RBF

- Use the automatic PCR procedure to choose the best approximation for each basis function.
- Consider all the approximations obtained previously as well as the approximations on the **x**-space. Remove the approximations that do not fulfill all the problem requirements, when the physical properties of the problem are used as criteria.

Table 1
Minimized CV errors for the data set with eight input variables

	Multiquadric CV error	Thin plate CV error	Cubic CV error	Gaussian CV error
$n = 8$	2697	18,814	3396	21,229
$r = 8$	4321	7113	3930	34,037
$r = 7$	3065	16,646	5505	7,175,000
$r = 6$	3724	4372	4157	615,560
$r = 5$	5837	4583	5745	29,257
$r = 4$	4163	5589	9100	7,757,400

Table 2
Relative errors for the leave-one-out CV of multiquadric PCR with $r = 7$

Airplane	Prediction	Relative error (%)	Airplane	Prediction	Relative error (%)
1049G	11,932	3.8	DC-7C	10,623	−4.5
C-46A	4274	−36.4	F-28	5046	−28.7
DC8F-54	34,963	2.5	B. 747	76,456	−14.1
DC9-30	11,764	−1.5	VC10-1101	33,986	0.2
VC10-1151	35,565	2.5	G-159	4399	18.8
C-8A	5958	30.9	AC-1	6151	59.5
C-124A	19,371	2.3	KC-97E	15,506	1.2
C-123J	5879	−2.0	XC-120	5327	−13.7
C-119H	11,894	−1.4	C. 440	8057	46.9
C. 340	4983	−6.9	R4Y-2	5327	6.4
C-131E	5174	3.9	Con.110	3200	−9.2
Con.T-29D	4147	−5.3	Martin 404	5118	4.6
V. V. 800	4878	−18.9	DC-6B	8492	4.6
B. 727	15,741	−12.7	CL44-D4-1	15,411	−2.9
C-130B	10,908	−2.2	Electra	8690	14.8
B. 720	24,341	3.6	Jetstar	5866	66.7
C5-A	77,204	−6.1	L-1011	35,511	−25.7
C-135A	29,220	15.1	22(880)	7280	5.6
30A(990A)	26,585	−0.1	C-141A	35,980	4.7
C-5A	82,296	−4.1	F-27	6150	36.2
DC10-10	48,536	0.1			

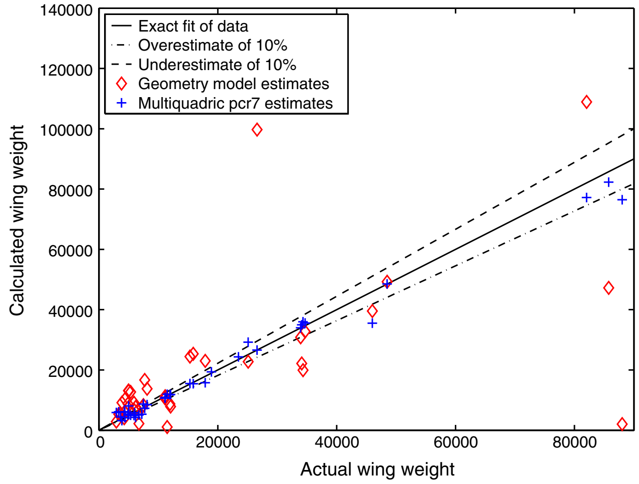

Fig. 4. Leave-one out CV errors for geometry model and multiquadric PCR with $r = 7$.

- Use the smallest sum of prediction errors

$$\sum_{j=1}^{N}(g_{-j}(\mathbf{x}^j) - f(\mathbf{x}^j))/g_{-j}(\mathbf{x}^j),$$

to choose the most adequate RBF and best approximation to use.

5. Concluding remarks

Radial basis function methods can provide excellent interpolants for a large number of poorly distributed data points. Regardless the methodology used, we should avoid a priori choice of RBF. Instead, a set of the most significant RBF should be considered and decide for the best RBF to use based on objective criteria. The numerical choice of the most adequate RBF proposed in the previous section can be an objective tool.

Among radial basis functions methods, Gaussian RBF methods, are one of the most referenced in literature for solving wing weight fitting problems. For our data set, Multiquadric RBF models presented the most desirable results. In fact, Gaussian RBF models presented poor results. This is another indication that the choice of the RBF to use should be part of the optimization problem instead of an a priori choice.

Principal component analysis, that identifies the collinear or nearly collinear variables, is a powerful tool that allows data fitting in reduced feature spaces. Cross-validation is useful for RBF interpolation model parameter tuning.

The significance of the wing weight problem study is that RBF interpolation methods (with PCA) are able to generate general approximation models for wing weight estimation that are accurate and have desirable properties. The choice of the RBF to use is decisive on the quality of the best approximation to use and on finding an approximation that improve the best existing approximations.

Even though the benefits of choosing the most adequate RBF to use and the corresponding best approximation are only demonstrated by the wing weight data fitting problem, the methodology could have significant advantages in fitting other historical or sparse data.

References

[1] H. Rocha, Principal Component Regression For Construction Of Wing Weight Estimation Models, Ph.D. Thesis, Old Dominion University, VA, USA, Department of Mathematics and Statistics, 2005.

[2] H. Rocha, W. Li, A. Hahn, Principal Component Regression Using Radial Basis Functions Interpolation, Wavelets and Splines: Athens 2005, Nashboro Press, Brentwood, 2006, pp. 402–415.

[3] H. Rocha, W. Li, A. Hahn, Principal component regression for fitting wing weight data of subsonic transports, J. Aircraft 43 (6) (2006) 1925–1936.

[4] M. Buhmann, Radial Basis Functions: Theory and Implementations, Cambridge University Press, Cambridge, UK, 2003.

[5] J. Carr, W. Fright, R. Beatson, Surface interpolation with radial basis functions for medical imaging, IEEE Trans. Med. Imaging (1997) 96–107.

[6] M. Powell, Radial basis function methods for interpolation to functions of many variables, HERMIS: Int. J. Comput. Maths Appl. 3 (2002) 1–23.

[7] M. Powell, Recent research at Cambridge on radial basis functions, new developments in approximation theory (Dortmund, 1998), Int. Ser. Numer. Math., Birkhauser, Basel, 1999, pp. 215–232.

[8] C. Baxter, The Interpolation Theory of Radial Basis Functions, Ph.D. Thesis, Cambridge University Department of Applied Mathematics & Theoretical Physics (DAMTP), August 2002.

[9] I. Schoenberg, Metric spaces and completely monotone functions, Ann. Math. (1938) 811–841.

[10] C. Micchelli, Interpolation of scattered data: distance matrices and conditionally positive definite functions, Construct. Approx. (1986) 11–22.

[11] R. Schaback, H. Wendland, Characterization and construction of radial basis functions, in: N. Dyn, D. Leviatan, D. Levin, A. Pinkus (Eds.), Multivariate Approximation and Applications, Cambridge University Press, 2001, pp. 1–24.

[12] S. Rippa, An algorithm for selecting a good value for the parameter c in radial basis function interpolation, Adv. Comput. Math. 11 (1999) 193–210.

[13] G.E. Fasshauer, J.G. Zhang, On choosing optimal shape parameters for RBF approximation, Numer. Algor. 45 (2007) 345–368.

[14] J. Tu, D.R. Jones, Variable Screening in metamodel design by cross-validated moving least squares method, in: 44th AIAA, 2003, Norfolk, Virginia.

[15] J. Tu, Cross-validated multivariate metamodeling methods for physics-based computer simulations, in: Proceedings of the IMAC-XXI, February 3–6, 2003, Kissimmee, Florida.

[16] M. Stone, Cross-validatory choice and assessment of statistical predictions, J. Royal Stat. Soc. 36 (1974) 111–147.

[17] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, 1993.

[18] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Am. Stat. Assoc. (2001) 1348–1360.

[19] R. Li, A. Sudjianto, Analysis of computer experiments using penalized likelihood Gaussian Kriging model, in: Proceedings of 2003 American Society of Mechanical Engineers (ASME) International Design Automation Conference, DETC2003/DAC–48758, 2003, Chicago, Illinois, USA.

[20] J. Fan, I. Gijbels, Local Polynomial Modeling and Its Applications, Chapman and Hall, London, 1996.

[21] G. Wahba, Spline Models for Observational Data, vol. 59, SIAM, 1990.

[22] M. Ardema, M. Chambers, A. Patron, A. Hahn, H. Miura, M. Moore, Analytical Fuselage and Wing Weight Estimation of Transport Aircraft, NASA Technical Memorandum 110392, May 1996.
[23] D. Raymer, Aircraft Design: A Conceptual Approach, third ed., AIAA, Reston, Virginia, 1999.
[24] W. Li, S. Padula, Approximation methods for conceptual design of complex systems, in: C. Chui, M. Neamtu, L. Schumaker (Eds.), Approximation Theory XI (Gatlinburg, 2004), Nashboro Press, Brentwood, TN, 2005, pp. 241–278.
[25] J. Nelder, R. Mead, A simplex method for function minimization, Comput. J. (1965) 308–313.