Examples of Compactly Supported Functions for Radial Basis Approximations

Arta A. Jamshidi and Michael J. Kirby
Department of Mathematics
Colorado State University,
Fort Collins, CO 80523,
e-mail:{jamshidi,kirby}@math.colostate.edu.

Abstract

Radial Basis Functions (RBFs) are widely used in science, engineering and finance for constructing nonlinear models of observed data. Most applications employ activation functions from a relatively small list, including Gaussians, multi-quadrics and thin plate splines. We introduce several new candidate compactly supported RBFs for approximating functions in $L^P(\mathbb{R}^d)$ via overdetermined least squares. We illustrate their utility on the benchmark Mackey-Glass time series data. We observe that these new RBFs significantly reduce the number of modes required to approximate the data and produce models that have significantly improved condition numbers.

Conference: The 2006 International Conference on Machine Learning; Models, Technologies and Applications (MLMTA'06)

Keywords: radial basis functions (RBFs), compact support, nonlinear function fitting, condition number

1. Introduction

The approximation of nonlinear relationships in scattered data is now a problem of established significance in science and engineering. In many instances physical considerations such as conservation laws can be exploited to understand, at least partially, the relationships between variables. However, often it occurs that only empirical observations of the phenomenon are available and relationships must be estimated by means of mathematical models such as neural networks or radial basis functions.

One of the main objectives in the construction of a model from known, or training, data is to optimize the quality of its performance on new data generated by the same process. Thus we require the models to have both descriptive and predictive features. While this goal can be approached from a variety of directions[1] the inherent conditioning of the model plays a critical role in its ability to generalize. In practice, if the data model is represented generally by the mapping

$$y = f(x)$$

we are concerned with how the output of the model changes as a consequence of perturbation of the input. In particular, if

$$y + \delta y = f(x + \delta x)$$

it is desirable that the magnitude of the change in the output $\|\delta y\|$ be small if $\|\delta x\|$ is small. By definition, well-conditioned models produce small variations in δy for small variations in δx.

For nonlinear mappings, such as those generated by multi-layer perceptrons, the estimation of the condition number is complicated by the fact that the Jacobian of the map must be estimated at every point of interest [9]. This is also true in general for radial basis functions. However, in the case of radial basis functions we can determine the the condition number associated with the perturbation of the parameters simply by computing the singular values of the interpolation matrix. This information provides an important measure of the sensitivity of the model.

In general the condition number of an $m \times n$ matrix is $O(mn)$ suggesting that (nonlinear) models that employ linear transformations have poor performance bounds for large data sets of sufficient complexity. We have observed however, that the nature of the condition number depends very significantly on the type of radial basis functions that are employed. With this motivation we considered several

[1] For example, regularization methods and cross validation.

forms of radial basis functions including those with compact support. We found that the functions generally available in the literature often have poor conditioning properties, at least for some of the data sets we have considered. In this paper we introduce several new compactly supported functions for approximating data that possess surprisingly good conditioning properties.

2. Radial Basis Functions for Approximating Scattered Data

A RBF expansion is a linear summation of qualifying nonlinear basis functions. In general, an RBF is a mapping $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ that is represented by

$$f(x) = Ax + \alpha_0 + \sum_{k=1}^{N_c} \alpha_k \phi(\|x - c_k\|_W) \qquad (1)$$

where x is an input pattern, ϕ is a radial basis function centered at location c_k, and α_k denotes the weight for kth RBF and A is an $m \times n$ matrix. The term W denotes the parameters in the weighted inner product

$$\|x\|_W = \sqrt{x^T W x}$$

The term $Ax + \alpha_0$ affords an affine transformation of the data and is useful so that the nonlinear terms are not attempting to fit flat regions. More general polynomials may be used for this purpose [13]. As usual, the dimensions of the input n and output m are specified by the dimensions of the input-output pairs from data.

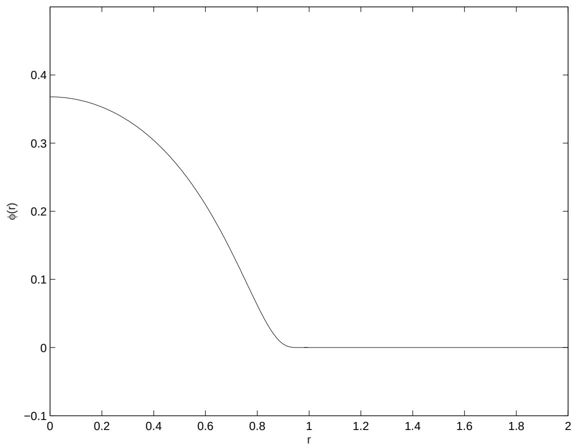
In this paper the implementation of the RBF follows our black-box methodology for nonlinear function approximation as described in [7]. It is assumed that the available data represents a functional relationship, or signal, with IID additive noise.

2.1. Compactly Supported RBFs

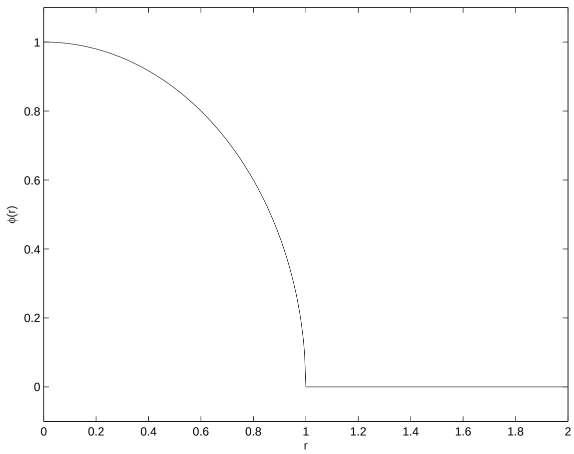
Recently several functions with compact support have been proposed as candidate radial basis functions, see, e.g., [15, 14, 16]. For example, the C^2 function

$$\phi(r) = (1 - r)_+^4 (1 + 4r) \qquad (2)$$

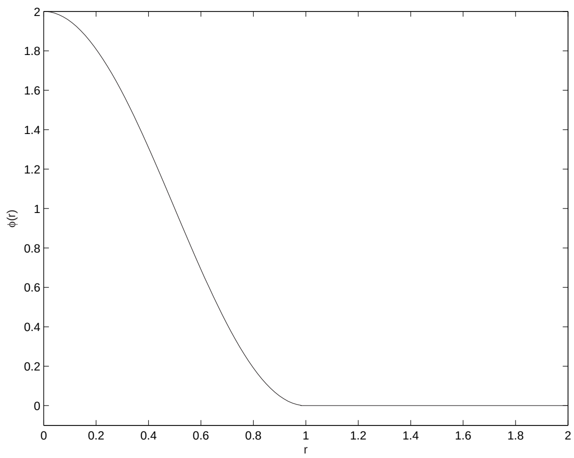
has been derived as an RBF explicitly for domain dimension 4 in the sense that the resulting square interpolation matrix is a (conditional) positive definite matrix [15]. In other words, we say that this function qualifies as an RBF given the square interpoliation matrix is guaranteed to be invertible. In many cases of practical interest it appears that this interpolation condition is overly restrictive. In general data fitting problem one is usually confronted with solving an overdetermined least squares problem. In this setting it seems adequate to require only that the approximating basis functions be dense in an appropriate function space.



(a) The plot of the C^∞ mollifier function.



(b) The plot of C^0 quarter circle.



(c) The plot of C^∞ Hanning function.

Figure 1. These functions can be used as $\phi(r)$ in the radial basis function expansion.

For example, as described in [10], the conditions required of basis functions to be dense in $L^P(\mathbb{R}^n)$ are very weak. For completeness, we briefly describe Park and Sandberg's theorem. Following [10], let K be a radially symmetric kernel function related to the activation function $\phi : [0, \infty) \longrightarrow \mathbb{R}$, such that, $K(\frac{x-c_i}{\sigma_i}) = \phi(\frac{\|x-c_i\|}{\sigma_i})$. The general element of the set $S_1(K)$ is expressed as

$$q(x) = \Sigma_{i=1}^{N_c} \alpha_i K(\frac{x - c_i}{\sigma_i}) \tag{3}$$

where $N_c \epsilon \mathbb{N}$, the set of natural numbers, is the number of basis functions, $\alpha_i \epsilon \mathbb{R}^m$ is the vector of weights, x is an input vector (an element of $\mathbb{R}^n$), c_i and σ_i are the center and widths of the ith kernel node, respectively. If $\sigma_i = \sigma$, i.e., all the widths are constant, then this family of functions is referred to as $S_0(K)$ [10].

Park and Sandberg's $S_0(K)$ Theorem [10]:

Let $K : \mathbb{R}^n \longrightarrow \mathbb{R}$ be an integrable bounded function such that K is continuous almost everywhere and $\int_{\mathbb{R}^n} K(x)dx \neq 0$. The family $S_0(K)$ is dense in $L^p(\mathbb{R}^n)$ for every $p \epsilon [1, \infty)$.

Park and Sandberg provide addition theorems for S_0 and S_1 with improved conditions in [11]. Motivated by these broad criteria which qualify functions as radial basis functions, we propose three different compactly supported functions that by Park and Sandberg's theorem are dense in $L^P(\mathbb{R}^n)$. In what follows we will illustrate their utility in practice in the context of over-determined least squares problem.

First, we propose the *bump* function widely used in differential geometry

$$\phi(r) = \exp(\frac{1}{r^2 - \gamma^2})H(1 - r^2)$$

for use as an RBF activation function where H is the usual Heaviside step function. This compactly supported and infinitely differential function is also widely referred to as a *mollifier*. It is shown in Figure 1 (a), and is qualitatively similar in nature to the widely applied non-compact Gaussian RBF, $\exp(-r^2)$. Interestingly, the failure of the Gaussian to have compact support has led some researches to arbitrarily truncate it. We observe that the Gaussian RBF satisfies the positive definiteness of the interpolation matrix for all space dimensions $d \geq 1$. Note that while the mollifier function satifies the postulates of Park and Sandberg's theorem, it has non-positive values in its Fourier transform and hence does not satisfy Wendland's *interpolation* criterion for a compact RBF [15]. Although this fact is of theoretical interest it is not of practical consequence since we are interested in the approximation (rather than interpolation) in the context of the overdetermined least squares problem.

A compact activation function with constant curvature is provided by

$$\phi(r) = \sqrt{1 - r^2}H(1 - r^2) \tag{4}$$

This is just the quarter circle shown in Figure 1 (b). Clearly this function also satisfies the postulates of Park and Sandberg's theorem. Of course this function is not differentiable where it meets the axis. While this could potentially cause problems in practice, we establish in the section on numerical experiments that the condition number of this RBF suggests it is worthy of further investigation.

Our last proposed activation function with compact support is the *Hanning* filter

$$\phi(r) = (\cos(r\pi) + 1)H(1 - r) \tag{5}$$

Like the bump function, this function is also infinitely differentiable; see Figure 1 (c). It has advantages over the mollifier function in the manner in which the function approaches zero, i.e., there is no vanishing term in a denominator. We do not present empirical results for this case in this current paper.

3. Numerical Examples

Here we employ a parsimonious growing algorithm with automatic mode determination as described in [6, 7]. This algorithm is an extension of the work in [1, 2] and is very attractive for comparing the qualities of various RBFs as it does not require any tuning of *ad hoc* parameters. The current algorithm works on higher dimensional domains and employs an spatio-temporal window to identify the data points that contribute to each RBF [7, 6]. The placement of the functions is driven by a statistical hypothesis test that reveals geometric structure when it fails. At each step the added function is fit to data contained in a spatio-temporally defined local region to determine the parameters, in particular, the scale of the local model.

3.1. Mackey-Glass Time Series

In this example we illustrate the mapping from a time-delay embedding of the univariate time-series to a future value. The Mackey-Glass time-delay equation

$$\frac{ds(t)}{dt} = -bs(t) + a\frac{s(t - \tau)}{1 + s(t - \tau)^{10}}. \tag{6}$$

generates a chaotic time series with short-range time coherence, where long time prediction is very difficult; it has become a standard benchmark for testing model fitting algorithms [12, 8, 17]. The time series is computed with parameters $a = 0.2$, $b = 0.1$ and $\tau = 17$ using the trapezoidal rule with $\Delta t = 1$, with initial conditions $x(t - \tau) = 0.3$ for $0 \leq t \leq \tau$ ($\tau = 17$). The initial 1000 data points corresponding to transient behavior are discarded. Then 4000 data points are reserved for the training set. The test set consists of 500 data points starting from point 5001. Note that not all 4000 training points collected were actually used for training the model. (These conditions are similar to those in Platt [12].)

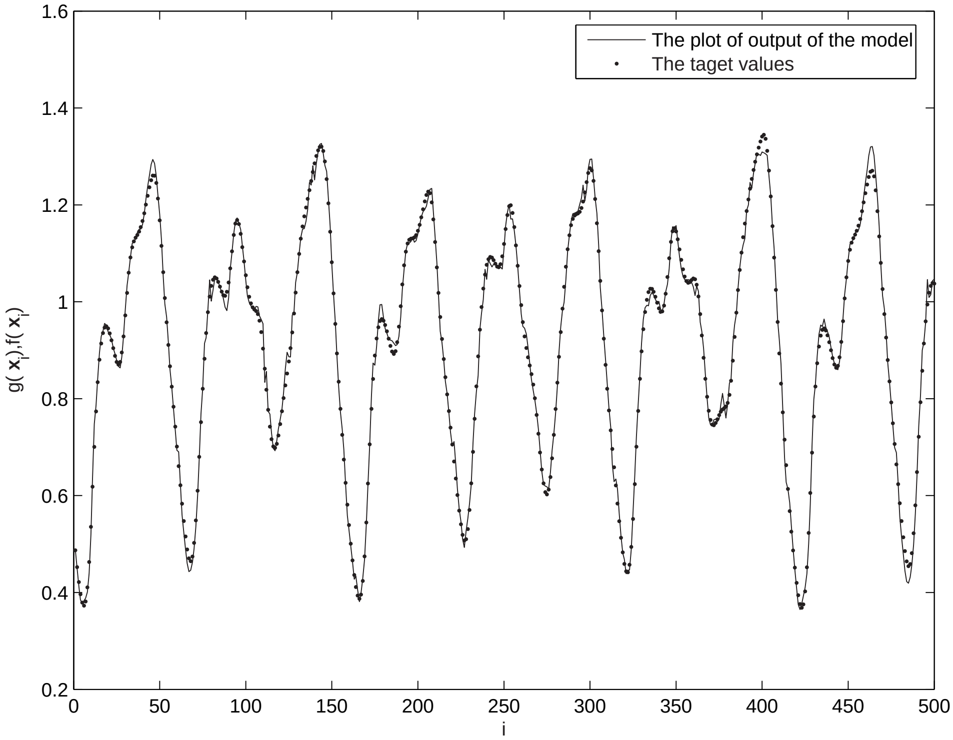


Figure 2. The output of the 37 mode model for the testing set compared to the target values. For this model an RMSE value of 0.0167 is obtained and the 95% of confidence stopping criteria was satisfied.

Table 1. This tables shows the performance of different RBFs under using an identical strategy of fit.

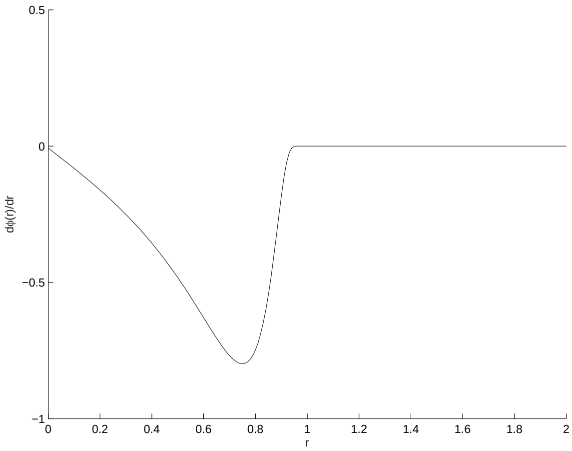
	Wendland RBF	Circle RBF	Mollifier
$ConditionNumber$	$3.0057e+003$	12.5845	284.3114
$RMSE$	0.0109	0.0344	0.0167
$NumberofRBFs$	51	26	37
$Confidence\%$	95	95.27	95.53

For purposes of comparison [18], the series is predicted with $\upsilon = 50$ samples ahead using four past samples: s_n, s_{n-6}, s_{n-12} and s_{n-18}. Hence, the nth input output data for the network to learn are

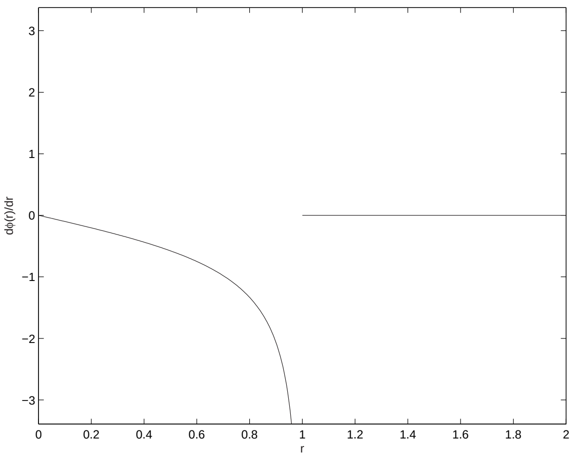
$$x_{n+\upsilon} = [s_n, s_{n-6}, s_{n-12}, s_{n-18}]^T$$

with $y_n = s_n$, whereas the υ step-ahead predicted value at time n is given by $z_{n+\upsilon} = f(x_{n+\upsilon})$, where $f(x_{n+\upsilon})$ is the network output at time n. The υ step-ahead prediction error is $\epsilon = s_{n+\upsilon} - z_{n+\upsilon}$. As such, this time series provides a good example for illustrating the construction of a mapping from $\mathbb{R}^4$ to $\mathbb{R}$. As a measure the performance of the various RBFs we compare the RMSE, the number of basis functions required and the sensitivity of the models via the condition number of the interpolation matrix of the full model. We present the final result of the fit using the mollifier in Figure 2. In this figure the output of the model and the associated target values are presented. The results of using other RBFs are summaried in Table 1. Note that all the results are aimed for 95% confidence in the statistical test [7, 6].

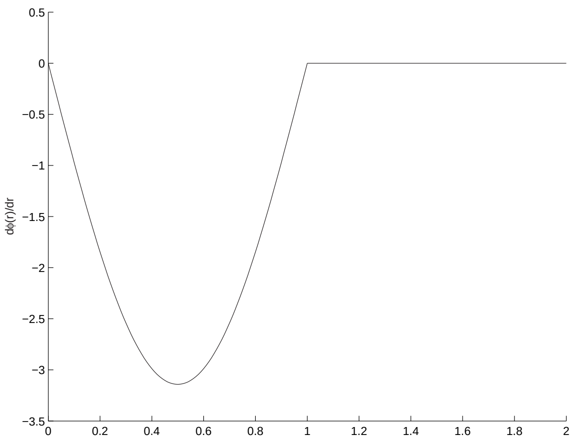
We make no claim that the RBFs proposed here are superior to other RBFs in the literature given there are so many factors that influence the performance of these fits. Certainly even the nature of the data set will dictate to some degree which RBF is appropriate. However it is significant that, on this data set at least, the condition numbers of the interpolation matrix are dramatically lower for the new RBFs. This could be due in part to the profile of the derivative of each of the RBFs. We see in Figure 3 (a) that the derivative of the mollifier is very small near the origin. The slope rises faster than that of the derivative of the quarter circle but is more well behaved for larger values. Obviously the circle suffers from the fact that it is not differentiable at $r = 1$ as shown in Figure 3 (b). Clearly this blow-up is potentially problematic but in our simulations the data for each RBF was never in this region. The symmetry of the Hanning derivative shown in Figure 3 (c) might have advantages but we also observe the steeper slope near the origin.



(a) The plot of the C^∞ mollifier function.



(b) The plot of C^0 quarter circle.



(c) The plot of C^∞ Hanning function.

Figure 3. The derivatives of the compact RBFs. Small values near $r = 0$ can lead to improved conditioning of the model.

4. Conclusions

We have proposed several new candidate compactly supported RBFs and have illustrated some of their positive performance properties on the benchmark Mackey Glass problem. Both the number of required modes and the conditioning of the final model are substantially improved over previous work. This suggests that RBFs proposed here provide additional options of interest in the data fitting problem. In particular, we advocate the use of either the mollifier function or Hanning RBF as an alternative to the truncated Gaussian RBF.

We note that the condition number of the interpolation matrix depends directly on the choice of RBF and suggested an explanation of the good conditioning properties of the proposed RBFs in terms of the behavior of the derivatives. It is interesting to speculate that new RBFs may be designed by optimizing the behavior of the derivative of the RBF for purposes of numerical conditioning. Such an approach should lead to improved function generalization.

The opportunity for applications of the proposed RBFs is significant, e.g., various problems in nonlinear signal processing, optimal control, computer vision, pattern recognition and prediction such as the financial time-series problem. In future work we will apply these functions for representing data on manifolds as graph of functions [4, 5] as well as the low-dimensional modeling of dynamical systems [3].

Acknowledgements: This study benefited from conversations with Gene Allgower and was supported by the National Science Foundation award DMS-0434351, DOD-USAF-Office of Scientific Research under contract FA9550-04-1-0094.

References

[1] M. Anderle. *Modeling Geometric Structure in Noisy Data*. Ph.D. dissertation, Colorado State University, Department of Mathematics, 2001.

[2] M. Anderle and M. Kirby. Correlation Feedback Resource Allocation RBFs. In *Proceedings of 2001 IEEE International Joint Conference on Neural Networks (IJCNN)*, volume 3, pages 1949–1953, Washington, D.C., July 2001.

[3] D. S. Broomhead and M. Kirby. Large Dimensionality Reduction Using Secant-based Projection Methods: The Induced Dynamics in Projected Systems. *To appear in Nonlinear Dynamics*.

[4] D. S. Broomhead and M. Kirby. A New Approach for Dimensionality Reduction: Theory and Algorithms. *SIAM Journal of Applied Mathematics*, 60(6):2114–2142, 2000.

[5] D. S. Broomhead and M. Kirby. The Whitney Reduction Network: a Method for Computing Autoassociative Graphs. *Neural Computation*, 13:2595–2616, 2001.

[6] A. A. Jamshidi. A New Spatio-Temporal Resource Allocation Network (ST-RAN). Master's thesis, Colorado State University, Fall, 2004.

[7] A. A. Jamshidi and M. J. Kirby. Towards a black box algorithm for nonlinear function approximation over high-dimensional domains. *To appear SIAM Journal of Scientific Computation*, in press.

[8] K. Kadirkamanathan and M. Niranjan. A Function Estimation Approach to Sequential Learning with Neural Networks. *Neural Computation*, 5(6):954–975, 1993.

[9] M. Kirby. Ill-conditioning and gradient based optimization of multi-layer perceptrons. In J.G. McWhirter and I.K. Proudler, editors, *Mathematics in Signal Processing IV*, The Institute of Mathematics and Its Applications Conference Series: No. 67, pages 223–237. Oxford University Press, 1998.

[10] J. Park and I. W. Sandberg. Universal Approximation Using Radial-Basis-Function Networks. *Neural Computation*, 3:246–257, 1991.

[11] J. Park and I. W. Sandberg. Approximation and Radial-Basis-Function Networks. *Neural Computation*, 5:305–316, 1993.

[12] J. Platt. A Resource Allocating Network for Function Interpolation. *Neural Computation*, 3:213–225, 1991.

[13] R. Schaback. *Reconstruction of Multivariate Functions from Scattered Data*. preprint, University of Göttingen, 1997.

[14] H. Wendland. Piecewise Polynomial, Positive Definite and Compactly Supported Radial Functions of Minimal Degree. *Advances in Computational Mathematics*, 4(1):389–396, December 1995.

[15] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2005.

[16] Z. Wu. Multivariate compactly supported positive definite radial functions. *Advances in Computational Mathematics*, 4:283–292, 1995.

[17] L. Yingwei, N. Sundararajan, and P. Saratchandran. A Sequential Scheme for Function Approximation Using Minimal Radial Basis Function Neural Networks. *Neural Computation*, 9:461–478, 1997.

[18] L. Yingwei, N. Sundararajan, and P. Saratchandran. Performance Evaluation of a Sequential Minimal Radial Basis Function (RBF) Neural Network Learning Algorithm. *IEEE Transactions on Neural Networks*, 9(2):308–318, March 1998.