Radial Basis Function Networks and Nonlinear Data Modelling.

Mats Carlin
SINTEF, Box 124 Blindern, N-0314 Oslo, Norway
Tel. +47 22 06 79 10, Fax. +47 22 06 73 50
Email: Mats.Carlin@ecy.sintef.no

1992

Abstract

Seven different Radial Basis Functions have been applied in a Feed-forward Neural Network and tested on five different real or simulated multivariate modelling problems.

A short theory of Radial Basis Functions is presented as well as the particular implementation of the Radial Basis Function Network (RBFN).

The real world data modelling problems are; identifying the dynamic actuator characteristics of a hydraulic industrial robot, modelling carbon consumption in a metallurgic industrial process and estimation of the water content in fish food products based on NIR-spectroscopy. In addition the RBFNs have been applied for modelling data generated from a simulated chemical reactor and to identify a 10-dimensional test function.

Key-words : Artificial Neural Networks, Radial Basis Functions, Nonlinear data modelling, Applications.

This paper has been published in **Proceedings of Neuro-Nimes'92**, Neural Networks and their Applications, EC2, France, 1992, pp.623-633.

1 A short theory of Radial Basis Functions.

Radial Basis Functions (RBFs) (or Potential Functions) are known from approximation theory as they are applied to the multivariate interpolation problem [Aizerman et.al. 64]. An RBF is generally described as

$$\Phi(r) = \Phi(\|\mathbf{x} - \mu\|); \mathbf{x} \in \Re^{\mathbf{n}}; \mathbf{r} \geq \mathbf{0}; \tag{1}$$

where $\Phi(r)$ is a continuous function on $(0, \infty)$ and its k'th derivatives are completely monotonic on $(0, \infty)$ for some k. $\|\cdot\|$ denotes a metric (usually the Euclidian norm or the Mahalanobis distance) and μ is the center (or mean) of the radial basis. Linear combinations of RBFs constitute the interpolation function :

$$y_i = \sum_j w_{ij} \Phi(\|\mathbf{x} - \mu_{\mathbf{j}}\|) \tag{2}$$

Radial Basis Functions can generally be divided into RBFs with local or global properties depending on the limiting value

$$\lim_{r \to \infty} \Phi(r) \tag{3}$$

The limiting value is zero for local RBFs and nonzero for global RBFs. Local RBFs interpolate only in a region of the input domain around its center, whereas global RBFs extrapolate globally. Early tests suggested that global RBFs actually are the best ones for interpolation problems in two dimensions [Franke 82].

This section will present some of the more common RBFs and their characteristics.

1.1 Gaussian Radial Basis Functions.

Most of the current research on Artificial Neural Networks with Radial Basis Functions are concentrating on using Gaussian RBF [Renals & Rohwer 89, Moody & Darken 89, Platt 91]. The general form is (fig.2) :

$$\Phi(r) = e^{-r^2} \tag{4}$$

described in [Powell 87, Broomhead & Lowe 88].

The Gaussian RBF has well-known mathematical features, it is highly nonlinear and provides good locality as a local RBF, which means that very efficient adaptive grid techniques can be used when implementing them on serial hardware [Moody & Darken 88, Moody & Darken 89]. Only small fractions of a large network responds to a particular input pattern. However this feature is not that important when realizing the neural network in parallel hardware.

Lee and Kil's Gaussian Potential Functions (GPF) are very similar to Radial Basis Functions in several manners, although they use a more complex distance metric including non-diagonal covariance elements [Lee & Kil 91].

Poggio and Girosi also states the fact that the Gaussian RBF is *the only Radial Basis Function which is factorizable* [Poggio & Girosi 89].

1.2 Pseudo-polynomial RBFs.

In the middle seventies Jean Duchon, proposed multi- conic and pseudo-cubic RBFs for the multivariate interpolation problem [Duchon 76].

The multi-conic (or linear) spline is described as :

$$\Phi(r) = r \tag{5}$$

whereas the pseudo-cubic spline (fig.3) is :

$$\Phi(r) = r^3 \tag{6}$$

These two RBFs are members of a larger family of RBFs [Duchon 76, Powell 87] given by :

$$\Phi(r) = r^{2i+1}, i \in \mathcal{N}_o \tag{7}$$

Together with "Thin Plate Splines" (TPS - section 1.4), these have an elegant mathematical theory in a Hilbert Space setting, minimizing a certain family of semi-norms in a Sobolev Space (which is a subspace of the Hilbert Space) [Duchon 76, Meinguet 79].

The pseudo-polynomial RBFs are all global RBFs.

1.3 MultiQuadric Equations (MQE).

Another method from the seventies, the MultiQuadric Equations (MQE), was proposed by Rolland Hardy and originally used for topographical mappings and surfaces [Hardy 71, Hardy 82].

The general form (fig.4) is :

$$\Phi(r) = \sqrt{r^2 + k^2} \tag{8}$$

which can be seen as taking the upper sheet of a circular hyperboloid of revolution. The parameter k must be specified by the user, but the method is quite stable with respect to this parameter and yields good results [Franke 82]. The MQE is a global RBF and the parameter k can be viewed as the smoothing factor of a multi-conic RBF as :

$$\lim_{k \to 0} \Phi(r) = r, r \geq 0 \tag{9}$$

Hardy also proposed a reciprocal (or inverse) MQE (fig.5) :

$$\Phi(r) = \frac{1}{\sqrt{r^2 + k^2}} \tag{10}$$

which is a local RBF and yields nearly as good results as the MQE itself [Franke 82], although it is more sensitive to the parameter k.

Micchelli has shown that the MQE is a special case of a larger family of RBFs [Michelli 86, Powell 87] :

$$\Phi(r) = (r^2 + k^2)^\beta, 0 < \beta < 1 \tag{11}$$

where the MQE has $\beta = \frac{1}{2}$.

The Inverse MQE (IMQE) is a special case of another large family of RBFs [Michelli 86, Powell 87] :

$$\Phi(r) = \frac{1}{(r^2 + k^2)^\alpha}, \alpha > 0 \tag{12}$$

where the IMQE has $\alpha = \frac{1}{2}$.

1.4 Thin Plate Splines (TPS).

As already mentioned, Duchon also proposed “Thin Plate Splines” (TPS) for the multivariate interpolation problem. Several other researchers had used them earlier, but Duchon and Meinguet developed the theory of the TPS [Duchon 76, Meinguet 79].

The general form of the TPS (fig.6) is :

$$\Phi(r) = r^2 \log r \tag{13}$$

which can be seen as minimizing the bending energy of a "Thin Plate" of infinite extent (thereby the name).

The TPS is a global RBF and performs well on the interpolation problem according to [Franke 82]. Duchon points out that the TPS is a member of the larger family of RBFs [Duchon 76, Michelli 86], given by :

$$\Phi(r) = r^{2i} \log r, i \in \mathcal{N}_o \tag{14}$$

1.5 The Logarithmic RBF.

Micchelli mentions another RBF used for the interpolation problem (fig.7) [Michelli 86] :

$$\Phi(r) = \log(r^2 + k^2) \tag{15}$$

which apparently has not been given any name. It will be referred to as the Logarithmic RBF (LRBF) throughout this paper. The LRBF is an interesting alternative to the other RBFs, as it has both local and global properties.

2 The Radial Basis Function Network (RBFN).

The RBFN is a Multilayer Feedforward Network with one hidden layer of Radial Basis Function Units (see figure 1).

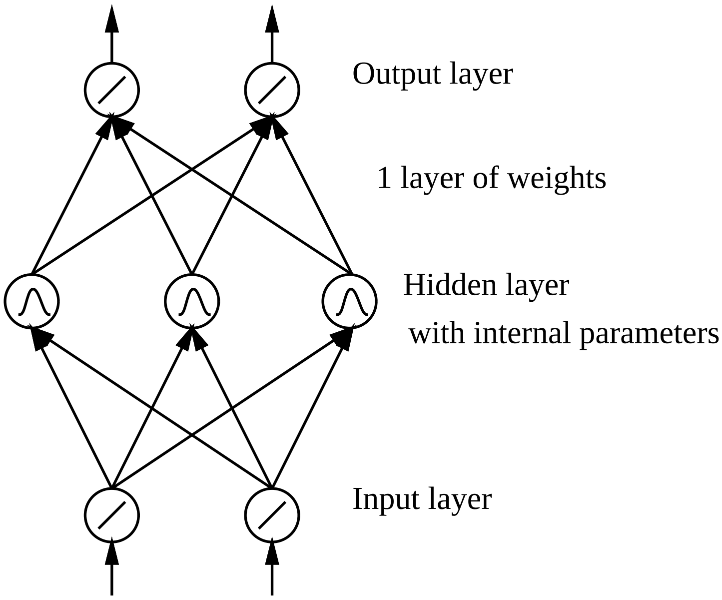

Figure 1: Multilayer Feedforward Network with Radial Basis Function.

The output layer performs a simple summation of the output from each hidden node :

$$y_j = \sum_i w_{ji} \Phi_i \tag{16}$$

where y_j is the j'th output node, w_{ji} is the weight from the i'th node of the hidden layer to the j'th node of the output layer and Φ_i is the output of the i'th hidden node. Φ_i describes the Radial Basis Function.

The following Radial Basis Functions have been implemented and tested:

- **Gaussian Radial Basis Function (GRBF)**
- **Inverse MultiQuadric Equation 1 (IMQE),** $\alpha = \frac{1}{2}$

- **Inverse MultiQuadric Equation 2 (ICUB),** $\alpha = 1$
- **Thin Plate Spline (TPS)**
- **Pseudo-Cubic Spline (PCUB)**
- **MultiQuadric Equation (MQE)**
- **Logarithmic Radial Basis Function (LRBF)**

The RBFN is implemented as described in [Carlin 91] using the Mahalanobis distance, which is defined as

$$r = \sqrt{\sum_l \frac{(x_l - \mu_l^i)^2}{\sigma_l^{i^2}}} \tag{17}$$

A modified version of the Hierarchical Self-Organized Learning (HSOL) [Lee & Kil 89] is used, with the following changes :

- Gaussian RBF : New nodes' weights are set to : $w_{jI} = t_j - y_j$.
- IMQE $\alpha = \frac{1}{2}$: New nodes' weights are set to : $w_{jI} = (t_j - y_j)k$.
- IMQE $\alpha = 1$: New nodes' weights are set to : $w_{jI} = (t_j - y_j)k^2$.

where t_j is the target and y_j the actual output of the network. All these changes provide better initial estimates. The HSOL algorithm automatically recruits the minimum necessary number of hidden nodes during learning for the given problem and initial parameters.

- For all global RBFs the initial weights are set to : $w_{jI} = 0$, since a non-zero initial weight would change the global network mapping.

Platt has proposed an almost similar algorithm called Resource-Allocating Network (RAN) [Platt 91].

All parameters $(\mu_l^i, \sigma_l^i, w_{ij})$ of the RBFN are updated by gradient descent. All parameter update rules can be found in [Carlin 91].

The only papers in my literature survey that presents results using Artificial Neural Networks with other RBFs than the Gaussian is [Broomhead & Lowe 88] using the MultiQuadric Equation and [Lowe 89] using the Thin Plate Spline.

3 Experimental results using the RBFN.

The Radial Basis Function Network has been applied to five different modelling tasks.

The Normalized Root Mean Square Error (NRMS) is used as a performance measure for the different models :

$$E_{NRMS} = \sqrt{\frac{\sum(t-y)^2}{\sum(t-\bar{t})^2}} = \sqrt{\frac{var(e)}{var(t)}} \tag{18}$$

which is the square root of the variance of the error over the variance of the target pattern.

All training sets and test sets are totally disjunct for all problems.

The RBFN has been tested on data taken from :

- A simulated catalytic chemical reactor.
- A 10-dimensional test function [Friedman 88].
- A metallurgic industrial process.
- A robot actuator.
- Water content in fish food products from NIR spectroscopy.

The simulated chemical reactor models a catalytic chemical process where unbranched hydrocarbons (nC_5) are transformed to branched hydrocarbons (iC_5) with a higher octane number. Hydrogen (H_2) is acting as a catalyst for the reactions. Crucial parameters are the temperature (T), the velocity of the process material through the catalytic tank (V) and the pressure (P). The different RBFNs have been used to predict iC_5 as

$$iC_5 = f(nC_{5_{init}}, iC_{5_{init}}, H_{2_{init}}, T, V) \tag{19}$$

where the data is generated by simulations of the catalytic process by simulations, based upon an analytical model in the form of differential equations. 700 randomly generated vectors are used as learning set, while 300 randomly generated vectors are used as test set.

The 10-dimensional test function is defined as

$$f(x) = 10\sin(\pi x_1 x_2) + 20(x_3 - .5)^2 + 10x_4 + 5x_5 + 0x_6 + \cdots + 0x_{10} \tag{20}$$

where the last five variables are totally uncorrelated with the output [Friedman 88]. White noise (20%) was added to the training data, but not on the test data. A set of 100 training points and 1000 test points were randomly taken from a uniform distribution over the 10 dimensional unit hypercube.

Elkem Sauda, a subsidary of Elkem a/s in Norway, produces FeMn alloy. At Sauda, the value of 75 process variables, averaged over 24-hour periods, were logged over a one year period. Our task was to model carbon consumption as a function of the 74 remaining variables. The number of variables were reduced using Partial Least Squares Regression. Three of the original variables were found to have a strong correlation with the carbon consumption and these three variables were used during modelling. Outlayers and incomplete data sets were removed and 13 different and totally disjunct test and training sets were used during 13 cross validation tests.

The RBFNs have been applied to identifying the characteristics of the servo valve/actuator system of the hydraulic TR4000 robot, from ABB Trallfa Robotics A/S. The goal was to find a general nonlinear function, $u(q, \dot{q}, \ddot{q})$, describing the required servo valve control signal, u, for a desired joint acceleration, $\ddot{q}$, and with given joint position, q, and velocity, $\dot{q}$. A more complete description of this experiment, using a B-spline modelling technique called ASMOD can be found in [Kavli 92].

Quantitative chemical analysis with NIR-spectrometry is one of the major application areas for Partial Least Squares Regression. Light absorbance or reflectance for a given light path through different samples of the analyte are measured at a number of frequencies. Since the different chemical components of the analyte have their characteristic absorption spectra it is possible to estimate the concentration of each component from the measured total absorbance spectrum. The problem in this experiment is to estimate water contents in fish food products, based on absorbance spectra measured at 19 frequencies in the NIR range. A set of 519 spectral measurements of test samples with known water contents was available, and 13 different sets of corresponding training and test sets were generated by randomly picking 50 of the measurements for each of the test sets.

All the test data are described in further detail in [Carlin et.al. 92] and compared with several other soft modelling schemes.

The result of applying the RBFNs these problems are shown in table 1. The results of using a linear model found by Partial Least Squares Regression are included for comparison.

Method	Chemical reactor	10-d function	FeMn process	Robot dynamics	NIR-spectra
GRBF	8.6%	32.8%	55.4(4.2)%	20%	21.6%
IMQE	7.9%	35.4%	55.0(5.3)%	19%	21.9%
ICUB	9.1%	37.6%	54.6(5.1)%	20%	23.2%
TPS	9.0%	-	56.1(7.7)%	33%	24.8%
PCUB	11%	-	-	76%	33.1%
MQE	6.5%	33.8%	53.6(6.7)%	31%	25.5%
LRBF	6.3%	27.9%	54.0(6.0)%	25%	22.8%
Linear	28%	-	58.4(6.6)%	64%	28.8%

Table 1: Comparison of all Radial Basis Functions and a linear model based on Partial Least Squares Regression shown for all data sets. The missing performance measures are due to lack of convergence during the tests.

The Radial Basis Function Networks were run on a SUN SPARC station with software developed at the Center for Industrial Research in Oslo. Typical learning times were in the range of ten minutes to a few hours, depending on the problem and the degree of refinement of the model defined by the number of hidden nodes.

4 Results and discussion.

The experiences gained using the different Radial Basis Functions can be summarized in the following :

- The RBFNs perform significantly better than the linear model for all test data with the exception of the 10-dimensional function, where the results should be compared with Friedman's MARS [Friedman 88], which gives a result of 19.5 % NRMS. The 10-dimensional test function is an example of a high dimensional problem with sparsely distributed data and relatively much noise on the training set. The Radial Basis Functions Network does not perform good under such conditions.

- The Logarithmic Radial Basis Function (LRBF) performs best on the two simulated examples and quite good on all the other examples.

- The Gaussian Radial Basis Function (GRBF) performs well on all examples. Both Inverse MultiQuadric Equations (IMQE and ICUB) are similar in form as the GRBF and show similar results.

- The MultiQuadric Equation (MQE) has variable performance on different test sets.

- The Thin Plate Spline is not very good at any of the problems, but its representation is more compact as it uses fewer hidden nodes.

- The Psedo-Cubic Spline performs poorly on all problems. There were difficulties of establishing convergence for the gradient search using this Radial Basis Function.

5 Summary.

Seven different Radial Basis Functions have been applied in a Multilayer Feedforward Network.

A short theory on Radial Basis Functions have been presented, as well as the specific implementation, called the Radial Basis Function Networks.

The Radial Basis Function Networks have been applied to five different nonlinear data modelling problems. The results indicate that the Logarithmic RBF have special properties which should be utilized in further work, while in general several of the Radial Basis Functions tested perform with near equal results.

Recommended reading.

It is hard to find good review articles on Radial Basis Functions, but the articles of [Powell 87, Poggio & Girosi 89] give thorough reviews. Applications of Multilayer Feedforward Networks using RBFs are described in [Broomhead & Lowe 88, Moody & Darken 89, Lee & Kil 91, Platt 91]. Only a few papers mention Radial Basis Functions applied to empirical modelling. Some important results are given in [Röscheisen et.al 91] on a very similar problem, predicting the rolling force of a rolling mill. Another approach worth mentioning is the Normalized Radial Basis Functions proposed by [Moody & Darken 88] and pursued by [Jones et.al. 90] and several other researchers.

Acknowledgements.

I would like to thank Tom Kavli and Bjørn Lillekjendlie at the Center for Industrial Research for valuable help, guidance and suggestions, as well as Alexander Øhrn for doing some of the computer programming.

This work was supported by a grant from The Royal Norwegian Council for Scientific and Industrial Research (NTNF).

References

[Aizerman et.al. 64] Aizerman, M.A., Braverman, F.M. & Rozoner, L.I. : *The Probability Problem in Pattern Recognition Learning and the Method of Potential Functions.* Automation and Remote Control 25, no.9, 1964.

[Broomhead & Lowe 88] Broomhead, D.S. and Lowe, David : *Multivariable Functional Interpolation and Adaptive Networks.* Complex Systems, vol.2, 1988, pp.321-355.

[Carlin 91] Carlin, Mats : *Neural Nets for Empirical Modelling.* Master Thesis, Norwegian Institute of Technology (NTH), 1991, pp.1-85.

[Carlin et.al. 92] Carlin, Mats et.al. : *A Comparison of Four Methods for Nonlinear Data Modelling.* To be published.

[Duchon 76] Duchon, Jean : *Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces.* In : Schempp & Zeller (eds) : *Constructive Theory of Functions of Several Variables.*, Lecture Notes in Mathematics, vol.571, Springer-Verlag, 1976, pp.85-100.

[Franke 82] Franke, R. : *Scattered Data Interpolation : Test of Some Methods.* Mathematics of Computation, vol.38, 1982, pp.181-200.

[Friedman 88] Friedman, Jerome : *Multivariate Adaptive Regression Splines (MARS).* Technical Report no.102, november 1988, Laboratory for Computational Statistics, Stanford University.

[Hardy 71] Hardy, Rolland L. : *Multiquadric Equations of Topography and Other Irregular Surfaces.* Journal of Geophysical Research, vol.71, 1971, pp.1905-1915.

[Hardy 82] Hardy, Rolland L. : *Surface Fitting with Biharmonic and Harmonic Models.* Proceedings of the

NASA workshop on Surface Fitting, Center for approximation theory, Texas A&M University, 1982, pp.136-146.

[Jones et.al. 90] Jones, R.D et al.: *Nonlinear Adaptive Networks: A Little Theory, a Few Applications.* Cognitive Modeling in Systems Control, Santa Fe, New Mexico, June 10-14, 1990.

[Kavli 90] Kavli, Tom : *Nonuniformly partitioned piecewise linear representation of continuous learned mappings.* Proceedings IEEE International Workshop on Intelligent Motion Control, 1990, pp.115-122.

[Kavli 92] Kavli, Tom : *ASMOD - An algorithm for Adaptive Spline Modelling of Observation Data.* To be published.

[Lapedes & Farber 88] Lapedes, Alan & Farber, Robert : *How Neural Nets Work.* In : Andersson(ed) : *Neural Information Processing Systems*, American Institue of Physics, 1988, pp.442-456.

[Lee & Kil 88] Lee, Sukhan & Kil, Rhee M. : *Multilayer Feedforward Potential Function Network.* IEEE International Conference on Neural Networks, San Diego : SOS Printing, vol.1, 1988, pp.161-171.

[Lee & Kil 89] Lee, Sukhan & Kil, Rhee M. : *Bidirectional Continuous Associator Based On Gaussian Potential Function Network.* International Joint Conference on Neural Networks, vol.1, Washington DC, June 18-22, 1989, pp.45-53.

[Lee & Kil 91] Lee, Sukhan & Kil, Rhee M. : *A Gaussian Potential Function Network With Hierarchically Self-Organizing Learning.* Neural Networks, vol.4, 1991, pp.207-224.

[Lowe 89] Lowe, David : *Adaptive Radial Basis Function Nonlinearities and The Problem of Generalization.*

IEE International Conference on Artificial Neural Networks, November 1989, pp.171-175.

[Meinguet 79] Meinguet, Jean : *Multivariate Interpolation at Arbitrary Points Made Simple.* Journal of Applied MAthematics and Physics (ZAMP), vol.30, 1979, pp.292-304.

[Michelli 86] Micchelli, Charles A. : *Interpolation of Scatteres Data : Distance Matrices and Conditionally Positive Definite Functions.* Constructive Approximation, vol.2, 1986, pp.11-22.

[Moody & Darken 88] Moody, John & Darken, Christian : *Learning with Localized Receptive Fields.* In : Touretzky et.al.(eds) : *Proceedings of the 1988 Connectionist Models Summer School.*, Morgan-Kaufman, 1988, pp.133-143.

[Moody & Darken 89] Moody, John & Darken, Christian : *Fast Learning in Networks of Locally-Tuned Processing Units.* Neural Computation, vol.1, 1989, pp.281-294.

[Platt 91] Platt, John : *A Resource-Allocating Network for Function Interpolation.* Neural Computation, vol.3, no.2, 1991, pp.213-225.

[Poggio & Girosi 89] Poggio, T. & Girosi, Federico : *A Theory of Networks for Approximation and Learning.* AI Memo no.1140, CBIP paper no.31, July 1989, MIT Press, pp.1-85.

[Powell 87] Powell, M.J.D. : *Radial Basis Functions for Multivariable Interpolation : A Review.* in Mason and Cox : *Algorithms for Approximation*, Clarendon Press, Oxford, 1987, pp.143-167.

[Renals & Rohwer 89] Renals, Steve & Rohwer, Richard : *Phoneme Classification Experiments Using Radial Basis Functions.* International Joint Conference on Neural

Networks, vol.1, Washington DC, June 18-22, 1989, pp.461-467.

[Röscheisen et.al 91] Röscheisen, Martin et.al. : *Incorporating Prior Knowledge in Parsimonious Networks of Locally-Tuned Units.* Technical Report FKI-155-91, Technical University of Munich, July 1991.

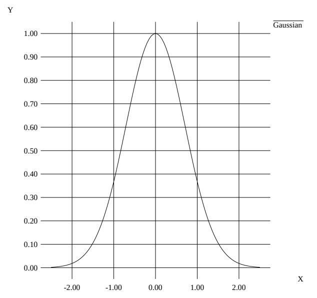

Figure 2: Gaussian Basis Function.

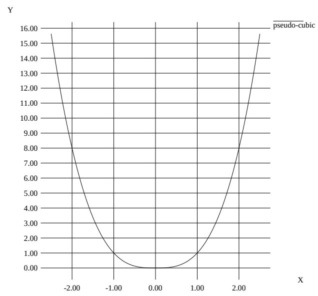

Figure 3: Pseudo-Cubic Basis Function.

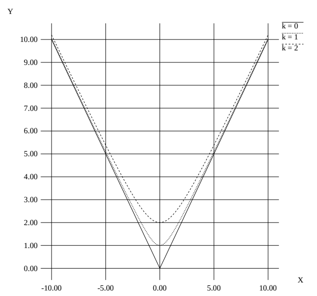

Figure 4: MultiQuadric Equations.

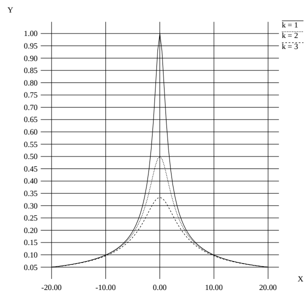

Figure 5: Inverse MultiQuadric Equations.

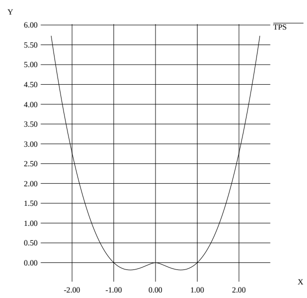

Figure 6: Thin Plate Spline.

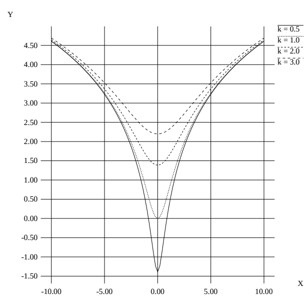

Figure 7: Logarithmic Basis Functions.