Radial Basis Functions and Their Applications

Xingping Sun
Missouri State University

August 1, 2006

What is a Radial Basis Function (RBF)?

In the Euclidean space $\mathbb{R}^n$ setting, an RBF is (roughly) a function of the form:

$$x \mapsto \sum_{j=1}^{N} c_j \phi(|x - x_j|),$$

where $x_1, \ldots, x_N$ are some scattered points, and $|x - x_j|$ denotes the Euclidean distance between x and x_j, $c_1, \ldots, c_N$ are some constants, and ϕ is a univariate function. In fact,

$$\phi : [0, \infty) \to \mathbb{R}.$$

ϕ is called the basis. Therefore, Radial Basis.

Example. Let $\phi(t) = \exp(-t^2)$. Then we have $\sum c_j \exp\left(|x - x_j|^2\right)$. This is called the Gaussian Radial Basis Function.

Why RBF?

Recall Lagrange interpolation, or spline interpolation in elementary numerical analysis. Data dealt with there are from $\mathbb{R}$. This is univariate interpolation. When the number of variables is big, and dimension is high, traditional methods become less efficient.

In Finite Element Methods, for example, a “mesh” or a “grid” is required before anything can be done. This is not practical, and becomes extremely expensive in high dimensional spaces.

Complexity grows exponentially with dimensions!

Radial Basis Functions or Radial Basic Functions?

William Light (Leicester, UK) advocated the latter.

Many data collected from real world problems have the following features:

1. Very stochastic!

2. Coming from spaces of high dimensions.

3. Scattered and noisy.

By using RBFs, we can establish many inexpensive probabilistic and deterministic error estimates for a variety of real world problems.

Fitting a surface to scattered data arising from sampling an unknown function defined on an underlying manifold comes up frequently in applied problems. When the underlying manifold is a sphere – or, more generally, the n-sphere $\mathbb{S}^n$ –, there are applications to geodesy, meteorology, astrophysics, geophysics, and other areas. Several review articles ([Fasshauer and Schumaker, 1998], [Mhaskar, Narcowich and Ward, 2001] and books ([Freeden, Gervens and Schreiner, 1998], [Wendland, 2005]) and a recent volume [**21** (2004)] of the journal *Advances in Computational Mathematics* have been devoted to the topic itself or its applications.

Currently, there are two main approaches to solving such problems. One can use spherical triangles and employ a local polynomial approximation. This approach is described in a review article by Fasshauer and Schumaker [1998], and recently Neamtu and Schumaker [2004] have derived error estimates for it.

Another approach, and the topic of this talk, is to use RBFs. These functions go back to work of Schoenberg [1937, 1938, 1942].

Advantages of RBF:

1. No "Curse of Dimensionality". Excellent performance in solving problems in high dimensions, such as problems in Machine Learning Theory.

2. Flexibility and robustness.

3. Less computational complexity.

The RBF Interpolation Scheme:

Given data $(x_j, d_j), j = 1, \ldots, N$, where $x_j \in \mathbb{R}^n$, and $d_j \in \mathbb{R}$, and a prescribed basis function ϕ, we intend to find the unknown coefficients $c_j, j = 1, \ldots,$ so that the radial basis function

$$\sum_{j=1}^{N} c_j \phi(|x - x_j|),$$

interpolates the data d_j. That is

$$\sum_{j=1}^{N} c_j \phi(|x_i - x_j|) = d_i, \quad i = 1, \ldots, N.$$

Is the problem well-posed?

This amounts to asking: Is the $N \times N$ matrix A with $ij-$entry $A_{ij} = \phi(|x_i - x_j|)$, nonsingular?

If ϕ is "strictly positive definite", or "strictly conditionally positive definite of order one", then the answer is yes; see Schoenberg [1938, 1942], Michelli [1986], Sun [1993] and the references therein.

The well-known Bochner's theorem on positive definite functions.

Definition 1. *A continuous function*

$$f : \mathbb{R}^n \to \mathbb{R}$$

is said to positive definite if for any N points $x_1, \ldots, x_N$ the matrix $\left(f(x_i - x_j\right)$ is positive definite.

Theorem 2. *(S. Bochner) A continuous function f is positive definite on $\mathbb{R}^n$ if and only if there is positive Borel measure μ on $\mathbb{R}^n$ such that*

$$f(x) = \int_{\mathbb{R}^n} e^{-i\langle x,\xi\rangle} d\mu(\xi).$$

In particular, if f has nonnegative and integrable Fourier transform, then f is positive definite. One can get many useful positive definite functions this way.

Proof of the Sufficiency of Bochner's Theorem.

$$\begin{aligned} & \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j f(x_i - x_j) \\ = & \int_{\mathbb{R}^n} |\sum_{j=1}^{N} c_j e^{i\langle \xi, x_j\rangle}|^2 d\mu(\xi) \\ \geq & \ 0. \end{aligned}$$

The proof of the necessity, however, is not very easy.

Some of of the most useful radial basis functions except the Gaussian.

$$\begin{aligned} \phi(t) &= t^{\alpha}, \quad 0 < \alpha < 2 \\ &= (1 + t^2)^{\alpha}, \quad \alpha < 0, \quad \text{or} \quad 0 < \alpha < 2. \end{aligned}$$

In particular, $(1 + t^2)^{1/2}$ and $(1 + t^2)^{-1/2}$ are called Hardy's multiquadrics and the inverse Hardy's multiquadrics, respectively, in honor of the Geologist Hardy who first employed the function to perform interpolation. If one uses the basis function $\phi(t) = t$, then the interpolation matrix is $(|x_i - x_j|)$, called the distance matrix.

How to show that the distance matrix ($|x_i - x_j|$) is non-singular? Using Fourier analysis technics, one can show that if $\sum_{j=1}^{N} c_j = 0$, and $\sum_{j=1}^{N} c_j^2 > 0$ then

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j |x_i - x_j| < 0.$$

Using the Min-max Theorem, we know that the distance matrix has $(N-1)$ negative eigenvalues. The trace of the matric is 0. So the matrix has at least one positive eigenvalue. Thus, the matrix has exactly one positive eigenvalue and $(N-1)$ negative ones.

Similar argument can used to handle other conditionally positive (negative) definite functions.

Just being non-singular is not enough. In the implementation of RBF interpolation, it is important to estimate the conditional number of the interpolation matrix A, which is a challenging task! Even a reasonable estimate of the smallest eigenvalue of the interpolation matrix needs some sophisticated Fourier analysis methods.

Another question, keen to mathematicians, is the following: how the interpolating RBF approximate the original (unknown) function when the data points become dense in the underlying domain? This is an active research area.

Regularization Scheme: Dealing with "Noisy Data".

Question: What are noisy data?

The Best Answer: All data are noisy in some ways.

Lets face the reality: when data are inevitably noisy, inaccurate, exact interpolation does not make sense, or even worse, it could be misleading! Thus, a regularization scheme is preferred in many real world problems.

Find an RBF from a suitable function class $\mathcal{H}_K$ so that the functional:

$$\sum_{j=1}^{N} \Big(f(x_j) - d_j \Big)^2 + \gamma \|f\|^2$$

is minimized over $f \in \mathcal{H}_K$, where $\mathcal{H}_K$ is a Reproducing Kernel Hilbert Space (RKHS) with reproducing kernel $K(x - y)$. The most useful reproducing kernels are again RBFs. This minimization scheme is advocated by Poggio, among others.

Theorem 3. *(Graven and Wabba [1979], Evgeniou, Pontil and Poggio [2000]) The function f_z that minimizes the regularized empirical error*

$$\frac{1}{N}\sum_{i=1}^{N}(d_i - f(x_i))^2 + \gamma\|f\|^2$$

over $f \in \mathcal{H}_K$, can be expressed as

$$f_z(x) = \sum_{i=1}^{N} c_i K(x - x_i),$$

where $c := (c_1, \ldots, c_N)$ is the unique solution of the well-posed linear system

$$(\gamma N \,\mathrm{Id} + K[x])c = d.$$

Here $K[x]$ denotes the interpolation matrix whose $ij-$entry is $K(x_i - x_j)$, and $d := (d_1, \ldots, d_N)$.

This theorem demonstrates two important facts:

1. RBFs provide optimal solutions to many problems.

2. Positive-definiteness bridges analysis (calculus of variation in this setting) and linear algebra.

Your Homework: Prove the above theorem.

Two kinds of estimates:

1. Probabilistic estimates. Data sites cannot be made uniform. Mathematical tools are mostly probabilistic inequalities, such as Chebysheff inequality, Bernstein inequality, Hoeffding inequality.

2. Deterministic estimates. If the distribution measure is known or mostly known, such as in some Monte Carlo methods, one can cater to the problem and select data sites, and get error estimates in terms of the minimal separations of the data sites.

Let $X = \{x_j\}_{j=1}^N \subset \mathbb{S}^n$ be a set of N distinct points on the sphere, and we will call X a set of centers. There are three useful quantities we will associate with X: the *separation radius*, q_X, the *mesh norm*, h_X, and the *mesh ratio*, ρ_X. If $d(x, y)$ is the geodesic distance between two points x and y in $\mathbb{S}^n$, then these quantities are defined by

$$\begin{aligned} q_X &:= \frac{1}{2} \min_{j \neq k} d(x_j, x_k), \\ h_X &:= \max_{x \in \mathbb{S}^n} \min_j d(x, x_j), \\ \rho_X &:= h_X / q_X. \end{aligned}$$

Note that $\rho_X \geq 1$.

On the circle $\mathbb{S}^1$, a set of N equiangular points has $q = h = \pi/N$, and $\rho = 1$. For $n > 1$, $\rho = 1$ cannot be achieved. Uniformly distributing a large number of points on spheres or other manifolds has been an extensively researched subject; see [Conway and sloan, 1993], [Habicht and waerden, 1951], [Hardin and Saff, 2004], [Saff and Kuijlaars, 1997].

Three Major Approaches: 1) Best Packings, 2)Spherical Designs, 3)Minimal Energies. For instance, Habicht and Van der Waerden [1951] studied the best packing of N non-overlapping hexagons on $\mathbb{S}^2$. A careful inspection of their proof shows that the X they constructed has a mesh ratio $\rho_X \leq 2/\sqrt{3} + CN^{-1/6}$, where C is a constant independent of N.

We sate without proof a simple algorithm of constructing arbitrarily large, nested sets of centers that also satisfy additional properties we need in the approximation scheme that happens at a later time.

Proposition 4. *There exists a sequence of sets* $X_k \in \mathcal{F}$, $k = 0, 1, \ldots$, *such that the sequence is nested,* $X_k \subset X_{k+1}$, *and such that at each step the mesh norms satisfy* $\frac{1}{4} h_{X_k} < h_{X_{k+1}} \leq \frac{1}{2} h_{X_k}$.

The proof does not involve any property special to $\mathbb{S}^n$, other than that it has a metric, finite volume, and a few other things associated with compact, connected C^∞ Riemannian manifolds. Thus it holds for these spaces.

Let $L^2(\mathbb{S}^n)$ be the Hilbert space equipped with the inner product

$$\langle f, g\rangle := \int_{\mathbb{S}^n} f(x)\overline{g(x)}d\mu(x),$$

where $d\mu$ is the standard volume element for $\mathbb{S}^n$, i.e., the restriction to $\mathbb{S}^n$ of the Lebesgue measure in $\mathbb{R}^{n+1}$. The $Y_{\ell,m}$'s will be taken to be the usual orthonormal basis of spherical harmonics. For ℓ fixed, these span the eigenspace of the Laplace-Beltrami operator on $\mathbb{S}^n$ corresponding to the eigenvalue $\lambda_\ell = \ell(\ell + n - 1)$. Here, $m = 1, \ldots, d_\ell$, where d_ℓ is the dimension of the eigenspace corresponding to λ_ℓ and is given by

$$d_\ell = \begin{cases} 1, & \ell = 0, \\ \dfrac{(2\ell + n - 1)\Gamma(\ell + n - 1)}{\Gamma(\ell + 1)\Gamma(n)}, & \ell \geq 1\,. \end{cases}$$