COS 702
Lecture 1: Scattered Data Interpolation

Globally Supported Radial Basis Functions

In many scientific disciplines one face the following problem. We get a set of data (measurements, and locations at which these measurements were obtained), and we want to find a rule which allows us to deduce information about the process we are studying also at locations different from those at which we obtained our measurements. Thus, we are trying to find a function which is "good" fit to the given data. There are many way to decide what we mean by "good", and the only criterion we will consider is that we want our approximation to exactly match the given measurements at the corresponding locations. This approach is called interpolation, and if the locations at which the measurement are taken do not lie on a uniform or regular grid, then the process is called *data interpolation.* More explicitly, we are considering the following

Problem 1 *Given data* $(\mathbf{x}_j, y_j), j = 1, \cdots, N$ *with* $\mathbf{x}_j \in \mathbb{R}^d, y_j \in \mathbb{R}$ *, find a (continuous) function* f *such that* $f(\mathbf{x}_j) = y_j, j = 1, \cdots, N$.

Here the $\mathbf{x}_j$ are the measurement locations (or data sites), and the y_j are the corresponding measurements (or data values). We will often assume that these values are obtained by sampling a data function f at the data sites, i.e., $y_j = f(\mathbf{x}_j), j = 1, \cdots, N$. The fact that we allow $\mathbf{x}_j$ to lie in d-dimensional space means that the formulation of Problem 1 allows us to cover many different types of problems. If $d = 1$, the data could be a series of measurements taken over a certain time period, thus the "data sites" $\mathbf{x}_j$ would correspond to certain time instances. For $d = 2$, we can think of the data being obtained over a planar region, and so $\mathbf{x}_j$ corresponds to the tow coordinates in the plane. For instance, we might want to produce a map which shows the rainfall in the state we live is based on the data collected at weather stations located through the state. For $d = 3$, we might think of a similar situation in space. One possibility is that we could be interested in the temperature distribution inside some solid body. Higher-dimensional problems might not be that intuitive, but a multitude of them exist, e.g., in finance, economics or statistics, but also in artificial intelligence or machine learning.

A convenient and common approach to solving the scattered data problem is to make the assumption that the function f is a linear combination of certain basis functions φ. Let $\{\mathbf{x}_i\}_{i=1}^n$ be distinct points on Ω. Let $\mathbf{x} = (x_1, x_2, \cdots, x_d)$ and $\varphi(\|\mathbf{x} - \mathbf{x}_j\|) : \mathbb{R}^d \to \mathbb{R}$, where $\|\cdot\|$ is the Euclidean norm, be a globally defined basis function. The theory of optimal recovery provides firm theoretical basis to assume $\hat{f}(\mathbf{x})$ in the form:

$$f(\mathbf{x}) \simeq \hat{f}(\mathbf{x}) = \sum_{j=1}^{n} a_j \varphi(\|\mathbf{x} - \mathbf{x}_j\|) \tag{1}$$

If the exactness condition,

$$f(\mathbf{x}_i) = \hat{f}(\mathbf{x}_i), \quad 1 \leq i \leq n,$$

is imposed then the linear system

$$f(\mathbf{x}_i) = \sum_{j=1}^{n} a_j \varphi(\|\mathbf{x}_i - \mathbf{x}_j\|), \quad 1 \le i \le n, \tag{2}$$

is uniquely solvable if the symmetric $n \times n$ matrix

$$\mathbf{A}_\varphi = \begin{bmatrix} \varphi(\|\mathbf{x}_1 - \mathbf{x}_1\|) & \cdots & \varphi(\|\mathbf{x}_1 - \mathbf{x}_n\|) \\ \vdots & \ddots & \vdots \\ \varphi(\|\mathbf{x}_n - \mathbf{x}_1\|) & \cdots & \varphi(\|\mathbf{x}_n - \mathbf{x}_n\|) \end{bmatrix} \tag{3}$$

is nonsingular. In the literature on RBFs, it was noted that for certain choices of RBFs (3) could be singular, at least for some configuration of interpolation points, and interest arose in finding sufficient conditions to ensure the existence of $\mathbf{A}_\varphi^{-1}$. It is well-known that positive definiteness is sufficient to guarantee the invertibility of the coefficient matrix $\mathbf{A}_\varphi$ in (3). However, most of the globally defined RBFs are only conditionally positive definite. In order to guarantee unique solvability of the interpolation problem, one adds a polynomial term to the interpolant (1) giving

$$\hat{f}(\mathbf{x}) = \sum_{j=1}^{n} a_j \varphi(\|\mathbf{x} - \mathbf{x}_j\|) + \sum_{k=1}^{t} b_k p_k(\mathbf{x}), \tag{4}$$

along with the constraints

$$\sum_{j=1}^{n} a_j p_k(\mathbf{x}_j) = 0, \quad 1 \le k \le t, \tag{5}$$

where $\{p_k\}_{k=1}^{t}$ is a basis for $\mathcal{P}_{m-1}$, the set of d-variate polynomials of degree $\le m-1$, and

$$t = \binom{m+d-1}{d}$$

is the dimension of $\mathcal{P}_{m-1}$. Let

$$\mathbf{P}^T = \begin{bmatrix} p_1(\mathbf{x}_1) & \cdots & p_1(\mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ p_t(\mathbf{x}_1) & \cdots & p_t(\mathbf{x}_n) \end{bmatrix}.$$

The interpolation conditions

$$f(\mathbf{x}_i) = \sum_{j=1}^{n} a_j \varphi(\|\mathbf{x}_i - \mathbf{x}_j\|) + \sum_{k=1}^{t} b_k p_k(\mathbf{x}_i), \quad 1 \le i \le n,$$

subject to the constraints in (5) can be rewritten as a linear system

$$\begin{bmatrix} \mathbf{A}_\varphi & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \\ \mathbf{0} \end{bmatrix}, \tag{6}$$

where $\mathbf{a} = [a_1, \cdots, a_n]^T, \mathbf{b} = [b_1, \cdots, b_t]^T$ and $\mathbf{F} = [f(\mathbf{x}_1), \cdots, f(\mathbf{x}_n)]^T$. This system of equations is uniquely solvable under the conditions that $\mathbf{A}_\varphi$ is nonsingular on the set of vectors $\mathbf{a}$ satisfying (5) and the polynomial in $\mathcal{P}_m$ should be uniquely determined by its values on $\{\mathbf{x}_i\}_{i=1}^n$.

For example in $\mathbb{R}^2$, if $\varphi(r) = r^2 \log r$, then it follows from Duchon's theorem that these conditions will be met if we choose $m = 1, t = 3$, $\{p_k\}_{k=1}^3 = \{1, x, y\}$,$\{\mathbf{x}_i\}_{i=1}^n$ are not collinear and then the system of $n + 3$ equations

$$\begin{cases} \sum_{j=1}^{n} a_j \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 \log \left\|\mathbf{x}_i - \mathbf{x}_j\right\| + b_1 + b_2 x_i + b_3 y_j = f(\mathbf{x}_i), \quad 1 \le i \le n, \\ \sum_{j=1}^{n} a_j = \sum_{j=1}^{n} a_j x_j = \sum_{j=1}^{n} a_j y_j = 0, \end{cases} \tag{7}$$

have a unique solution.

COMPACTLY SUPPORTED RBFS

During the past two decades radial basis functions have been intensively studied and widely applied to many areas of science. On the other hand, the mathematical analysis and construction of compactly supported positive definite rbfs (CS-PD-RBFs) have only recently been established.
The degree of the polynomial function from Wendland turns out to be minimum in terms of the given order of smoothness. Recently, CS-PD-RBFs have also been applied to multivariate surface reconstruction. Here we use CS-PD-RBFs as a tool to approximate the forcing term f.

We first give some notation and a brief review of CS-PD-RBFs. We assume the radius of support to be normalized to $[0, 1]$.

Definition 2 *A continuous function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is positive definite iff there exists $\Phi : \mathbb{R}_+ \to \mathbb{R}$ such that $\varphi(\mathbf{x}) = \Phi(\|\mathbf{x}\|)$ where $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^d$ and the quadratic form*

$$\lambda^T \mathbf{A}_\varphi \lambda = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \varphi\left(\|\mathbf{x}_i - \mathbf{x}_j\|\right), \quad \lambda = (\lambda_1, \cdots, \lambda_n),$$

is strictly positive when the vector $\lambda \neq 0$ and all set of distinct points in $\mathbb{R}^d$. (In this case it is customary to say that φ is positive definite as well[8].) Note that if $\varphi \in \mathbf{PD}_d$ then $\varphi \in \mathbf{PD}_k$ whenever $k \leq d$.

In other words, if φ is $\mathbf{PD}_d$, then the matrix $\mathbf{A}_\varphi$ in (3) induced by this function will be positive definite. Therefore, the system of equations (2) has a unique solution. Stable numerical methods such as the Cholesky factorization are available for symmetric positive definite matrices. We also define the following piecewise polynomial

$$(1-r)_+^n = \begin{cases} (1-r)^n, & \text{if} \quad 0 \leq r \leq 1, \\ 0, & \text{if} \quad r > 1. \end{cases}$$

The following theorem gives the existence and eventually leads to the construction of the CS-PD-RBFs.

Theorem 3 *(Wendland) Up to a constant factor there exists exactly one positive definite function φ of the form*

$$\varphi(r) = \begin{cases} p(r), & \text{if} \quad 0 \leq r \leq 1, \\ 0, & \text{if} \quad r > 1. \end{cases}$$

where $p(r)$ is a univariable polynomial with $\deg p(r) = \lfloor d/2 \rfloor + 3k + 1$ and $\varphi \in C^{2k}$. Here $\lfloor d/2 \rfloor$ stands for the integer n satisfying $n \leq d/2 < n + 1$.

CS-PD-RBFs have also been constructed explicitly by a recursive formula as shown in Wendland's paper. As a result, for any given dimension d and smoothness C^{2k}, a minimum degree polynomial function $\varphi(r)$ can always be constructed. A polynomial with minimum degree is always preferable for numerical implementation. A list of Wendland's function is given in Table 1. The CS-PD-RBF $\varphi = (1-r)_+^4(4r+1)$ in Figure 1 possesses two smooth derivatives around zero and three smooth derivatives around 1.

Table 1. Wendland's CS-PD-RBFs.

$d=1$	$\varphi=(1-r)_+$	C^0
	$\varphi=(1-r)_+^3(3r+1)$	C^2
	$\varphi=(1-r)_+^5(8r^2+5r+1)$	C^4
$d=3$	$\varphi=(1-r)_+^2$	C^0
	$\varphi=(1-r)_+^4(4r+1)$	C^2
	$\varphi=(1-r)_+^6(35r^2+18r+3)$	C^4
	$\varphi=(1-r)_+^8(32r^3+25r^2+8r+1)$	C^6

Since the supports of all these function have been normalized to $[0,1]$, we can rescale functions in Table 1 with support of radius α by using $\varphi(r/\alpha)$ for $\alpha>0$.

Let us introduce a scaling parameter $\alpha>0$ and denote $\varphi_\alpha(r)=\varphi(r/\alpha)$. It is easy to see that the matrix $\mathbf{A}_\varphi$ in (3) is diagonal when α is sufficiently small. The system of equations (2) is trivially solvable using $O(n)$ operations. In this case, $\hat{f}$ in (1) is a linear combination of sharp spikes and the quality of $\hat{f}$ will be poor. With the increase of α so does the density of $\mathbf{A}_\varphi$. This means that more information about the interpolation at each data center is provided and thus improvements in the quality of $\hat{f}$ and $\hat{u}_p$ are expected. Meanwhile, $\mathbf{A}_\varphi^{-1}$ becomes more difficult to evaluate and computational efficiency decreases. As α becomes large enough to cover the whole domain, φ turns out to be a global basis function which is what we are trying to avoid in this paper. Eventually, one has to compromise the quality of approximation and computational efficiency by choosing a proper scaling factor α according to the requirements of the particular application.