

RBF interpolation and Gaussian process regression through an RKHS formulation

Ken Anjyo and J. P. Lewis

Received on February 11, 2011

Abstract.

Radial Basis Function (RBF) interpolation is a common approach to scattered data interpolation. Gaussian Process regression is also a common approach to estimating statistical data. Both techniques play a central role, for example, in statistical or machine learning, and recently they have been increasingly applied in other fields such as computer graphics. In this survey we describe the formulation of both techniques as instances of functional regression in a Reproducing Kernel Hilbert Space. We then show that the RBF and Gaussian Process techniques can in some cases be reduced to an identical formulation, differing primarily in their assumptions on when the data locations and values are known, as well as in their (respectively) deterministic and stochastic perspectives. The scope and effectiveness of the RBF and Gaussian process techniques are illustrated through several applications in computer graphics.

Keywords. Reproducing Kernel Hilbert Space, Radial Basis Function, scattered data interpolation, Gaussian process regression,

1. INTRODUCTION

Radial Basis Function (RBF) interpolation is a common approach to scattered data interpolation. Gaussian Process regression (GPR) can also be used for scattered data interpolation and extrapolation. In this paper we interpret both of these techniques in the common framework of Reproducing Kernel Hilbert Spaces (RKHSs). The RKHS interpretation helps reveal the commonalities of these two techniques as well as clarify their differences. The practical relevance of the techniques will also be illustrated through several examples in computer graphics.

RBF Interpolation was popularized in the late 1980s, initially in the machine learning community [4]. More recently RBF interpolation has had numerous applications in other fields. For example, in computer graphics RBF interpolation has been used for reconstructing models from scanned data [5], for defining the skin motion of characters [10, 8], for interpolating volume medical scans [15], for editing lighting and shading on anime characters [17], and other purposes.

Gaussian process regression was initially explored by Kolmogorov and Wiener in the 1940s. It was later developed in the geostatistics community where it goes by the name Kriging [6]. Gaussian process techniques have been widely applied, for example in computer graphics Gaussian processes have been used for modeling the motion of human characters [7, 14], for synthesizing terrains [9] and for generating novel variations of human drawings [2].

Both RBF interpolation and Gaussian process regression involve the inverse problem of estimating an unknown function from a finite number of data samples. We describe how an RKHS is the appropriate theoretical basis for both techniques. In particular, for our practical purposes of solving interpolation/regularization problems in an RKHS, we utilize a powerful feature of RKHSs, known as the Representer Theorem [16]. This theorem provides effective solutions to function estimation problems in the form of weighted combinations of a kernel function. Given finite data samples, seeking an unknown function of this simple form is much easier to justify than a search within the unrestricted (infinite-dimensional) Hilbert space. Appropriate choices of the kernel function can summarize the known or assumed characteristics of the particular class of functions, thus leading to sensible regularization of the inverse problem. The RBF and GPR techniques can be seen as reflecting different approaches to specifying these assumptions.

In addition, linear operations within the RKHS can be interpreted as nonlinear operations in the input space (this is known as the “kernel trick” in the machine learning community). This enables problems to be solved as linear systems while still having the some of the power of a nonlinear approach.

The RBF and GPR techniques and the notion of RKHS are defined in the next section. The RKHS formulation of RBF interpolation is presented in section 3. Section 4 derives GPR from functional regression in an RKHS.

Section 5 applies the preceding discussion to establish an equivalence between RBF interpolation and GPR.

2. DEFINITIONS AND THEORY

2.1. REPRODUCING KERNEL

Let E be an abstract set, and \mathbf{H} be a Hilbert space consisting of the (real-valued¹) functions defined on E , with the inner product $\langle \cdot, \cdot \rangle$.

Definition 1. The function $K : E \times E \rightarrow \mathbb{R}$ is called a *reproducing kernel* of \mathbf{H} , if it satisfies the following conditions²:

1. For any fixed $y \in E$, $K(x, y)$ belongs to \mathbf{H} as a function of x on E .
2. For any $f \in \mathbf{H}$, we have $f(y) = \langle f(x), K(x, y) \rangle_x$.

Definition 2. If Hilbert space \mathbf{H} has the above kernel K , then \mathbf{H} is referred to as a *reproducing kernel Hilbert space* (RKHS).

The following proposition will be used in characterizing the reproducing kernel in the next section.

Proposition 1. For the reproducing kernel K , we have:

$$(1) \quad K(y, z) = \langle K(x, y), K(x, z) \rangle_x$$

Proof. From condition 1 of the reproducing kernel, we have $K(y, z) \in \mathbf{H}$ for a fixed z . Then, by putting $f(y) = K(y, z)$ in condition 2, we have

$$K(y, z) = \langle K(x, z), K(x, y) \rangle_x = \langle K(x, y), K(x, z) \rangle_x.$$

The next theorem is well known as a classical result in the reproducing kernel theory (N. Aronszajn [1], S. Bergman [3]).

Theorem 1. *Hilbert space $\mathbf{H}(E)$ has a reproducing kernel, if and only if the following condition is satisfied:*

For any $y \in E$, there exists a positive constant $C = C_y$, such that

$$(2) \quad |f(y)| \leq C_y \|f\|, \text{ for any } f \in \mathbf{H}.$$

Proof. [only if part] The assertion follows from Schwarz' inequality. [if part] For a fixed $y \in E$, let us consider the linear functional $\delta_y : \mathbf{H}(E) \rightarrow \mathbb{R}$, which is defined as $\delta_y(f) := f(y)$, for $f \in \mathbf{H}(E)$. Condition (2) then means that δ_y is continuous. The assertion thus follows from Riesz' Theorem [21]. That is, for a fixed $y \in E$, there exists a unique function $K_y(x) \in \mathbf{H}(E)$ such that $\delta_y(f) = \langle f, K_y \rangle$ for any $f \in \mathbf{H}(E)$. Then we put $K(x, y) = K_y(x)$.

Relating to the above theorem, we note that the reproducing kernel K is uniquely determined for an RKHS $\mathbf{H}(E)$ (also see Theorem 3 in the next section).

¹This article treats only real-valued functions, while the extension to complex-valued functions is obvious.

²In condition 2, the inner product $\langle \cdot, \cdot \rangle_x$ means that we get the inner product value of the two functions with variable x .

RKHS Examples

1. \mathbb{R} : Let E be $\{1\}$. Here we identify $\mathbf{H}(E)$ with \mathbb{R} . An arbitrary element of $\mathbf{H}(E)$ is a map $f : E \equiv \{1\} \rightarrow \mathbb{R}$. Specifying $f \in \mathbf{H}(E)$ therefore means specifying a real number x with $f(1) = x$. Let the inner product $\langle \cdot, \cdot \rangle$ for $\mathbf{H}(E)$ be the ordinary multiplication in $\mathbb{R} : \langle x, y \rangle = x \cdot y$. We define $K : E \times E \rightarrow \mathbb{R}$ as $K(1, 1) := 1$. It is then easy to see K satisfies condition 1 in the definition of the reproducing kernel. As for condition 2, we have: $\langle f(1), K(1, 1) \rangle = f(1)1 = f(1)$.
2. l^2 : Let $\mathbf{a} = (a_j)_{j=1}^\infty \in l^2$. Then \mathbf{a} defines a map $\alpha : \mathbb{N} \rightarrow \mathbb{R}$ with $\alpha(i) := a_i (i \in \mathbb{N})$. We thus identify $\mathbf{a} \leftrightarrow \alpha$. By setting $E = \mathbb{N}$, we have $\mathbf{H}(\mathbb{N}) \equiv \{\alpha : \mathbb{N} \rightarrow \mathbb{R} \mid \sum_{i=1}^\infty |\alpha(i)|^2 < \infty\} \cong l^2$ with its kernel function K as being $K(i, j) = \delta_{ij}$.
3. Let A be an n -th order, symmetric, and positive semi-definite matrix. Then $A(\mathbb{R}^n)$ is RKHS and its reproducing kernel is A (see the discussions in the next section):

$$A(\mathbb{R}^n) \equiv \{Ax \in \mathbb{R}^n \mid x \in \mathbb{R}^n\}.$$

2.2. FUNDAMENTAL PROPERTIES

Theorem 2. *Let $K : E \times E \rightarrow \mathbb{R}$ be the kernel function of RKHS $\mathbf{H}(E)$. Then K satisfies the following properties:*

1. *K is symmetric: $K(x, y) = K(y, x)$ for any $x, y \in E$.*
2. *K is positive semi-definite: For any $(x_1, \dots, x_n)^T \in E^n$ and $(a_1, a_2, \dots, a_n)^T \in \mathbb{R}^n$, we always have $\sum_{i,j=1}^n a_i a_j K(x_i, x_j) \geq 0$.*

Proof. (1) in Proposition 1 says that K is symmetric, since the inner product itself is symmetric. For $(x_1, \dots, x_n)^T \in E^n$ and $(a_1, a_2, \dots, a_n)^T \in \mathbb{R}^n$, it follows from (1) that

$$\begin{aligned} \sum_{i,j=1}^n a_i a_j K(x_i, x_j) &= \sum_{i,j=1}^n a_i a_j \langle K(x, x_i), K(x, x_j) \rangle_x \\ &= \left\langle \sum_{i=1}^n a_i K(x, x_i), \sum_{j=1}^n a_j K(x, x_j) \right\rangle_x \\ &= \left\| \sum_{k=1}^n a_k K(x, x_k) \right\|_x^2 \geq 0. \end{aligned}$$

The following theorem says that RKHS is constructed by specifying a symmetric, positive semi-definite function. It should also be noted that the proof is constructive so that it might be useful even in our practical situations.

Theorem 3. *Suppose that K is a symmetric, positive semi-definite function on $E \times E$. Then there exists a Hilbert space \mathbf{H} that has K as its reproducing kernel.*

Sketch of the proof. We put $\mathbf{F} := \{\sum_{i=1}^l \alpha_i K(x, x_i) \mid l \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_i \in E\}$. By defining addition and multiplication by constant as usual, we can make \mathbf{F} a vector space.

Also we can introduce the inner product for the expressions $f = \sum_{i=1}^m \alpha_i K(x, x_i)$ and $g = \sum_{j=1}^n \beta_j K(x, x_j) \in \mathbf{F}$ as follows: $\langle f, g \rangle := \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j K(x_i, x_j) \in \mathbb{R}$. Since K is positive semi-definite, it follows that $\langle f, f \rangle \geq 0$. It is easy to see that \mathbf{F} is a pre-Hilbert space. Next we set \mathbf{H} as the completion³ of \mathbf{F} . Then, with $g(x) \equiv g_y(x) = K(x, y)$, we have $\langle (f(x), K(x, y))_x = \langle f, g_y \rangle = \sum_{i=1}^m \alpha_i K(x_i, y) = \sum_{i=1}^m \alpha_i K(y, x_i) = f(y)$, for any $f \in \mathbf{F}$. This also holds for any f of \mathbf{H} , because, for any Cauchy sequence $\{f_m\}$ of \mathbf{F} , we have

$$(3) \quad \begin{aligned} |f_m(y) - f_n(y)| &\leq | \langle (f_m - f_n)(x), K(x, y) \rangle_x | \\ &\leq \|f_m - f_n\|_x \cdot \|K(x, y)\|_x \\ &= \|f_m - f_n\| \cdot \sqrt{K(y, y)}. \end{aligned}$$

This means that $\{f_m(y)\} \subset \mathbb{R}$ converges at any y . \mathbf{H} therefore contains $f = \lim_{n \rightarrow \infty} f_n$, because f also satisfies condition 2 in the definition of reproducing kernel.

One more remark is about an RKHS that has the complete orthonormal system (CONS). Let us describe the reproducing kernel K with $\text{CONS} \equiv \{\varphi_j\}_{j=1}^\infty$. From condition 1 of the reproducing kernel, we first get

$$K(x, y) = \sum_{i=1}^\infty \alpha_i(y) \varphi_i(x).$$

Then, taking $f(y) = \varphi_i(y)$ in condition 2, we have

$$\begin{aligned} \varphi_i(y) &= \langle \varphi_i, K(\cdot, y) \rangle = \langle \varphi_i, \sum_{k=1}^\infty \alpha_k(y) \varphi_k \rangle \\ &= \sum_{k=1}^\infty \alpha_k(y) \langle \varphi_i, \varphi_k \rangle = \sum_{k=1}^\infty \alpha_k(y) \delta_{ik} = \alpha_i(y). \end{aligned}$$

We therefore have

$$K(x, y) = \sum_{i=1}^\infty \varphi_i(x) \varphi_i(y).$$

2.3. RKHS IN L^2 SPACE

We briefly describe an infinite-dimensional RKHS in $L^2(E)$, where E is assumed to be a domain in \mathbb{R}^n . We refer to L^2 instead of $L^2(E)$ for simplicity. The inner product $(\cdot, \cdot)_{L^2}$ in L^2 is then defined as: $(f, g)_{L^2} = \int_E f(x)g(x)dx$ for any $f, g \in L^2(E)$. Supposing that K is a symmetric, positive semi-definite function on $E \times E$, we first define the real-valued function $\kappa(f)$ on E : $\kappa(f)(y) := \int_E K(x, y)f(x)dx$, for any $y \in E$. Let us further suppose that

$$(4) \quad \iint_{E \times E} |K(x, x')|^2 dx dx' < \infty.$$

³The completion simply means the space that consists of all the limits of Cauchy sequences of \mathbf{F} .

Then κ can be considered as a linear operator: $L^2 \rightarrow L^2$, because, using Schwarz' inequality, we have

$$\begin{aligned} \int_E |\kappa(f)(y)|^2 dy &= \int_E \left(\int_E K(x, x')f(x)dx \right)^2 dx' \\ &\leq \int_E \left(\int_E |K(x, x')|^2 dx \int_E |f(y)|^2 dy \right) dx' \\ &= \iint_{E \times E} |K(x, x')|^2 dx dx' \int_E |f(y)|^2 dy \\ &= \iint_{E \times E} |K(x, x')|^2 dx dx' \cdot \|f\|_{L^2}^2 < \infty. \end{aligned}$$

This yields that $\kappa(f) \in L^2$ and that κ is a continuous linear operator, known as *Hilbert-Schmidt integral operator*. According to Mercer's theorem, we then have the eigen decomposition: $K(x, x') = \sum_{\nu \geq 1} \lambda_\nu \phi_\nu(x) \phi_\nu(x')$ where $\nu \in \mathbb{N}$, and λ_ν, ϕ_ν are eigen value and eigen functions of κ , respectively. Assumption (4) yields $\sum_{\nu \geq 1} \lambda_\nu^2 < \infty$, so that we have $\lim_{k \rightarrow \infty} \lambda_k = 0$. We now assume that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq \dots > 0$. We then know that $\{\phi_n\}_{n=1}^\infty$ is a CONS of L^2 , which consequently gives the following result:

Theorem 4. *Let \mathbf{H}_κ be the totality of the functions $f \in L^2$, satisfying $\sum_{k \geq 1} \frac{f_k^2}{\lambda_k} < \infty$, where $f_k = (f, \phi_k)_{L^2}$, referred to as the Fourier coefficient of f w.r.t. $\{\phi_n\}_{n=1}^\infty$. We then have:*

1. \mathbf{H}_κ is a Hilbert space with the inner product: $\langle f, g \rangle_\kappa = \sum_{k \geq 1} \frac{f_k g_k}{\lambda_k} < \infty$, where $f, g \in \mathbf{H}_\kappa$, with $f_k = (f, \phi_k)_{L^2}$ and $g_k = (g, \phi_k)_{L^2}$.
2. For $g \in \mathbf{H}_\kappa$, we have

$$\langle g, \phi_\nu \rangle_\kappa = \frac{(g, \phi_\nu)_{L^2}}{\lambda_\nu},$$

$$\langle \phi_\nu, \phi_\nu \rangle_\kappa = \frac{1}{\lambda_\nu} \text{ (i.e., } \|\phi_\nu\|_\kappa = \frac{1}{\sqrt{\lambda_\nu}} \text{)} \quad (\nu = 1, 2, \dots).$$
3. K is the reproducing kernel of \mathbf{H}_κ : $f(x) = \langle f, K(\cdot, x) \rangle_\kappa$ for any $f \in \mathbf{H}_\kappa$.

We skipped the mathematical details and the rigorous proof of the above theorem. Instead, we should keep in mind the relation between \mathbf{H}_κ and L^2 through the CONS derived from the Hilbert-Schmidt operator κ . We also note that a similar result is obtained in a finite-dimensional case, where K simply means an n -th order symmetric, positive semi-definite matrix and $L^2 \cong \mathbb{R}^n$.

2.4. RADIAL BASIS FUNCTIONS

The form

$$(5) \quad \hat{f}(\mathbf{x}) = \sum_k^N w_k G(\mathbf{x} - \mathbf{x}_k)$$

is an example of radial basis interpolation of a function $f(\mathbf{x})$ known at a set of points \mathbf{x}_k . The function $G(\cdot)$ is assumed to be a radially symmetric function of its argument, i.e. $G(r) := \phi(|r|)$.

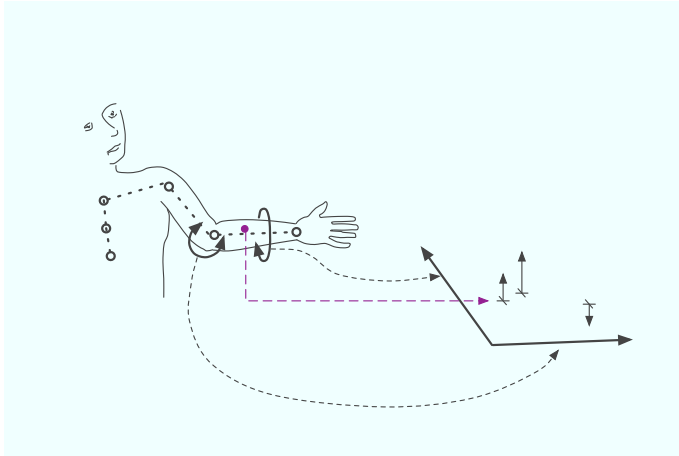


Figure 1: Schematic diagram of RBF interpolation for pose space deformation [10]. The coefficients of each vertex are interpolated from example shapes in different poses. In the figure, the position of a particular vertex on the fore-arm is interpolated as a function of several examples situated in a “pose space” described by the elbow and fore-arm rotations.



Figure 2: An animated human hand created with WPSD (reproduced from [8]). After obtaining example hand poses from medical scans, the resulting WPSD model can produce a realistic hand in any pose.

Given known values $y_k \equiv f(\mathbf{x}_k)$, the weights w_k can be solved as a linear system

$$(6) \quad \begin{bmatrix} G(\mathbf{x}_1 - \mathbf{x}_1) & G(\mathbf{x}_1 - \mathbf{x}_2) & G(\mathbf{x}_1 - \mathbf{x}_3) & \cdots \\ G(\mathbf{x}_2 - \mathbf{x}_1) & G(\mathbf{x}_2 - \mathbf{x}_2) & \cdots & \\ G(\mathbf{x}_3 - \mathbf{x}_2) & \cdots & & \\ \vdots & & & \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}$$

Since the argument to the kernel function $G()$ depends only on the distance between points, the interpolation takes the form of a weighted sum of radially symmetric functions.

Common choices of the kernel function $G()$ with $\phi(r)$ are:

- Gaussian $\phi(r) = \exp(-(r/c)^2)$
- Hardy multiquadratic $\phi(r) = \sqrt{r^2 + c^2}, c > 0$

Definition 3. [19] A function $G : \mathbb{R}^n \rightarrow \mathbb{R}$ is *positive semi-definite* if, for any $N \in \mathbb{N}$, all sets of N distinct points $\mathbf{x}_k \in \mathbb{R}^n$, and any $\mathbf{w} = (w_1, w_2, \dots, w_N)^T \in \mathbb{R}^N$, the quadratic form $\sum_{i=1}^N \sum_{j=1}^N w_i w_j G(\mathbf{x}_i - \mathbf{x}_j)$ is non-negative. The function G is called *positive definite* if the quadratic form is strictly positive, for any $\mathbf{w} \neq 0 \in \mathbb{R}^N$.

The linear system (6) of course has a solution for a positive definite $G()$.

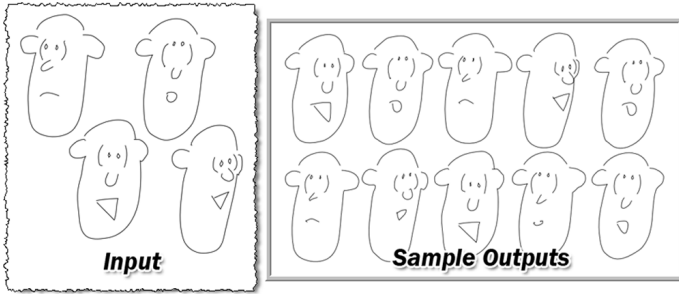
In other cases the matrix $G_{i,j} \equiv G(\mathbf{x}_i - \mathbf{x}_j)$ has a significant nullspace. A prominent example is spline interpolation of scattered data [13, 18], to be discussed further in section 3. In the spline case the interpolation has the form

$$(7) \quad \hat{f}(\mathbf{x}) = \sum_k^N w_k G(\mathbf{x} - \mathbf{x}_k) + p_{m-1}(\mathbf{x})$$

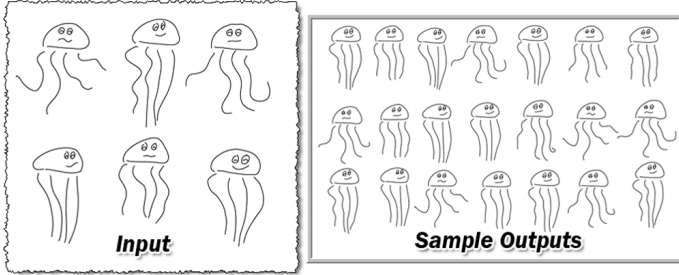
where $p_{m-1}(\mathbf{x}) \in \mathcal{P}_{m-1}$. Here we denote the space of polynomials of degree at most k by \mathcal{P}_k .

RBF Application Examples

1. Pose space deformation (PSD)[10]: This technique is an example-based skinning method. The word “skinning” refers to the skin (surface) deformation of a 3D character such as human as the underlying skeleton moves (Figure 1). PSD uses skin surface examples of varying posture and interpolates them during animation as a function of the vector of joint angles of the skeleton. The skin is prescribed with its vertices, so that RBF interpolation is applied to the coefficients of those vertices. In the notation of (5), \mathbf{x} is the vector of joint angles or other parameters describing the current pose of the skeleton, \mathbf{x}_k is the pose of the k th example, and $\hat{f}(\mathbf{x})$ is the interpolated value of one of the vertex coefficients. The weights w_k are different for each vertex coefficient. Kurihara and Miyata [8] introduced the improved weighted PSD (WPSD) algorithm. In this approach the distance between poses $\|\mathbf{x}_j - \mathbf{x}_k\|$ is different at each vertex, with individual terms in the distance being weighted by the influence of that parameter on the particular vertex. This effectively replaces the single pose space in PSD with smaller coupled pose spaces at each vertex, thereby greatly reducing the number of needed training examples. Both PSD and WPSD use RBFs of Gaussian type (and its normalized version in [8]).
2. Learning doodles by example: Baxter and Anjyo [2] proposed the concept of a latent doodle space, a low-dimensional space derived from a set of input doodles, or simple line drawings in \mathbb{R}^2 . This approach first gives a heuristic algorithm for finding stroke correspondences between the input drawings, and then proposes a few latent variable techniques to automatically extract a low-dimensional latent doodle space from the inputs. By the stroke correspondence algorithm, each of the input drawings is represented as a feature vector by combining all the x and y coordinates of each point on each stroke into one vector. One of the latent variable techniques first employs PCA (Principal Component Analysis) applied to the feature vectors. The first two principal components are then used to constitute the 2-dimensional latent doodle space, and, using thin plate spline RBF, we can obtain new drawings in this 2-d space (Figure 3)



(a) cartoon face



(b) jellyfish drawing

Figure 3: “Doodle” synthesis. The left images were drawn by an artist; the right images were then synthesized using the technique in [2].

2.5. GAUSSIAN PROCESS REGRESSION

A Gaussian process is a function from which any finite number of samples have a jointly Gaussian distribution. A Gaussian process $u(\mathbf{x})$ is thus completely specified by its mean $\mathbf{E}[u(\mathbf{x})]$ and covariance function $\Sigma_u(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{E}[u(\mathbf{x}_1)u(\mathbf{x}_2)]$. Gaussian process regression (GPR) estimates the values of a Gaussian process given observed values at other locations.

A value $u(\mathbf{x})$ of a Gaussian process at location \mathbf{x} can be estimated as

$$(8) \quad \hat{u}(\mathbf{x}) = \boldsymbol{\sigma}^T \boldsymbol{\Sigma}_u^{-1} \mathbf{y}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the vector of known data values, and $\boldsymbol{\sigma} = \mathbf{E}[u(\mathbf{x}), u(\mathbf{x}_k)]$ is the covariance between $u(\mathbf{x})$ (the function at the point being evaluated) and the function at the data points \mathbf{x}_k . If it is assumed that the data \mathbf{y} are observed in the presence of an independent identically distributed Gaussian observation noise ϵ then (8) takes the form

$$(9) \quad \hat{u}(\mathbf{x}) = \boldsymbol{\sigma}^T (\boldsymbol{\Sigma}_u + \lambda \mathbf{I})^{-1} \mathbf{y}$$

with λ being the variance of the observation noise.

Expression (8) is derived by assuming a linear estimate

$$(10) \quad \hat{u}(\mathbf{x}) = \mathbf{w}^T \mathbf{y}$$

The weights \mathbf{w} for the optimal linear estimate are found by applying the orthogonality principle, i.e., the estimate

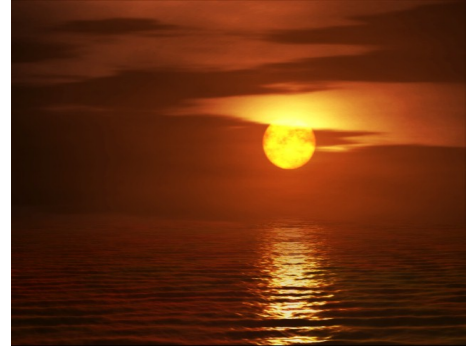


Figure 4: Synthetic ocean waves generated by Gaussian process regression with a specified covariance [9].

is optimal when the error is orthogonal in expectation to the data:

$$\mathbf{E}[(\hat{u}(\mathbf{x}) - u(\mathbf{x})) y_k] = 0 \quad \forall k$$

Although the true value $u(\mathbf{x})$ is unknown, substituting (10) and applying the expectations $\mathbf{E}[y_j y_k] = \Sigma(\mathbf{x}_j, \mathbf{x}_k)$ and $\mathbf{E}[u(\mathbf{x}) y_k] = \Sigma(\mathbf{x}, \mathbf{x}_k)$ results in (8) with $\mathbf{w} = \boldsymbol{\Sigma}_u^{-1} \boldsymbol{\sigma}$. Note that the linear estimate (10) is in fact the optimal estimator (among all linear and non-linear estimators) if the underlying function is in fact Gaussian [20].

Another useful expression is the variance of the regression estimate. This can be obtained from $\mathbf{E}[(\hat{u}(\mathbf{x}) - u(\mathbf{x}))^2]$ by substituting (10), resulting in

$$(11) \quad \mathbf{E}[(\hat{u}(\mathbf{x}) - u(\mathbf{x}))^2] = \mathbf{E}[u(\mathbf{x})^2] - \boldsymbol{\sigma}^T \boldsymbol{\Sigma}_u^{-1} \boldsymbol{\sigma}$$

GPR Application Examples

1. Kriging:

Gaussian process regression is form of Kriging, a term used in geostatistics to refer to estimation of quantities based on spatial samples and covariance (or variogram) information [11]. More specifically, (8) is identical to *simple Kriging*, whereas the more commonly used *ordinary Kriging* adds an additional constraint that the weights \mathbf{w} sum to one. In addition to its use in geostatistical interpolation, Kriging has recently been applied in computer graphics for interpolating human body poses from motion capture [14].

2. Terrain Synthesis:

Gaussian process regression is also called Wiener interpolation in the signal estimation literature. [9] utilized this interpolation technique in a hierarchical subdivision scheme to incrementally synthesize random textures and terrains. Whereas the fractal approach to terrain synthesis is restricted to the $1/f^p$ family of power spectra, in Wiener interpolation an arbitrary covariance is specified, allowing the synthesis of non-fractal landscapes such as ocean waves (Figure 4). Note that the covariance function and power spectrum are a Fourier transform pair by the

Wiener-Khinchine theorem. The goal in texture and terrain synthesis is to obtain a representative sample from the underlying random process. To achieve this, in the synthesis each variable is first estimated with (10), but then this estimate is displaced by an i.i.d. (independent identically distributed) Gaussian pseudo-random variable with variance (11).

3. RKHS AND RADIAL BASIS FUNCTIONS

3.1. REGULARIZATION PROBLEM IN RKHS

For simplicity, let $E = \mathbb{R}^n$. Suppose that (\mathbf{x}_i, y_i) are given as sample points, where $y_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^n (i = 1, 2, \dots, N)$. Let us consider the following regularization problem: Find a function f defined on \mathbb{R}^n such that

$$(12) \quad \min_f \left\{ \sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2 + \lambda J_m^n(f) \right\},$$

where

$$(13) \quad J_m^n(f) := \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_n = m} \frac{m!}{\alpha_1! \alpha_2! \dots \alpha_n!} \|D^\alpha f\|_{L^2}^2,$$

$$D^\alpha f := \frac{\partial^m f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}.$$

The regularization term $\lambda J_m^n(f)$ in (12) prescribes smoothness of a solution. Where can we find a solution of this problem (12)? According to the definition of $J_m^n(f)$, we should find a solution to (12) in the following space:

$$(14) \quad \mathbf{B}_m^n := \{f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\} \mid D^\alpha f \in L^2(\mathbb{R}^n), \text{ for any } \alpha (|\alpha| = m)\}.$$

Now recall the thin plate spline case. We then start with \mathbf{B}_2^2 in (14) to minimize the energy functional

$$(15) \quad F(\varphi) := \iint_{\mathbb{R}^2} \left(\left| \frac{\partial^2 \varphi}{\partial x_1^2} \right|^2 + 2 \left| \frac{\partial^2 \varphi}{\partial x_1 \partial x_2} \right|^2 + \left| \frac{\partial^2 \varphi}{\partial x_2^2} \right|^2 \right) dx_1 dx_2.$$

In the following we want to solve the regularization problem like (12) in RKHS. The main reason for this is the following nice property of RKHS:

Representer Theorem⁴ Let \mathbf{H} be an RKHS, with its reproducing kernel K and norm $\|\cdot\|_{\mathbf{H}}$. Consider the regularization problem of the following form: Find $f \in \mathbf{H}$ such that

$$(16) \quad \min_{f \in \mathbf{H}} \left\{ \sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2 + \lambda \|f\|_{\mathbf{H}}^2 \right\}.$$

The solution f can then be found in the form:

$$(17) \quad f(\mathbf{x}) = \sum_{i=1}^N \alpha_i K(\mathbf{x}, \mathbf{x}_i).$$

⁴This is one of the variations of the representer theorem. Please refer to [16].

It would therefore be nice, if we could have \mathbf{B}_m^n as the RKHS in the above theorem. However, J_m^n cannot be the squared norm for \mathbf{B}_m^n , as described next.

The functional J_m^n has the following properties:

1. $J_m^n(f) = 0 \Leftrightarrow f \in \mathcal{P}_{m-1}$.
2. $J_m^n(f) = (-1)^m \langle f, \Delta^m f \rangle_{L^2}$, where Δ is Laplacian: $\Delta := \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$.

Since the null space of J_m^n is equal to \mathcal{P}_{m-1} , we first represent \mathbf{B}_m^n as being the direct sum of the two function spaces: $\mathbf{B}_m^n = \mathbf{H}_m^n \oplus \mathcal{P}_{m-1}$. Then let us solve the regularization problem on \mathbf{H}_m^n . As a result of [12], we therefore have:

Theorem 5. If $m > \frac{n}{2}$, then \mathbf{H}_m^n is an RKHS with:

$$(18) \quad \langle f, g \rangle_{\mathbf{H}_m^n} := \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_n = m} \frac{m!}{\alpha_1! \alpha_2! \dots \alpha_n!} \langle D^\alpha f, D^\alpha g \rangle_{L^2} = \langle (-1)^m \Delta^m f, g \rangle_{L^2}.$$

This also means $\|f\|_{\mathbf{H}_m^n}^2 = J_m^n(f)$.

With the above theorem, the regularization problem (12) is restated as:

$$(19) \quad \min_{f \in \mathbf{B}_m^n} \left\{ \sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2 + \lambda J_m^n(f) \right\}$$

$$= \min_{g \in \mathbf{H}_m^n, p \in \mathcal{P}_{m-1}} \left\{ \sum_{i=1}^N \{y_i - (g(\mathbf{x}_i) + p(\mathbf{x}_i))\}^2 + \lambda \|g\|_{\mathbf{H}_m^n}^2 \right\}.$$

We thus know the solution to (19) is represented in the form of

$$(20) \quad f(\mathbf{x}) = \sum_{i=1}^N \alpha_i K(\mathbf{x}, \mathbf{x}_i) + p_{m-1}(\mathbf{x}),$$

where $p_{m-1}(\mathbf{x}) \in \mathcal{P}_{m-1}$. Here we denote the space of polynomials of degree at most k by \mathcal{P}_k .

3.2. RBF AS GREEN'S FUNCTION

To describe RBF in the above RKHS framework, we first consider the function $K_{\mathbf{y}}(\mathbf{x}) = K(\mathbf{x}, \mathbf{y})$ ⁵, where $K(\mathbf{x}, \mathbf{y})$ is the kernel of \mathbf{H}_m^n . In [12], it is shown that $K_{\mathbf{y}}$ gives a weak solution⁶ of a certain type of non-homogeneous linear differential equation of m -th iterated Laplacian Δ^m . Next let G be a Green's function in the sense that

$$(21) \quad \Delta^m G(\mathbf{x}) = \delta(\mathbf{x}),$$

⁵The function $K_{\mathbf{y}}(\mathbf{x})$ is the same as $K_{\mathbf{y}}(x)$ appeared in the proof of Theorem 1.

⁶The weak solution is a solution in the sense of distributions [21].

where δ is the Dirac delta (generalized) function. Then it is further shown in [12] that we can replace equation (20) by

$$(22) \quad f(\mathbf{x}) = \sum_{i=1}^N \alpha_i G(\mathbf{x} - \mathbf{x}_i) + p(\mathbf{x}),$$

where p is a polynomial $\in \mathcal{P}_{m-1}$. This brings us to our familiar class of radial basis functions:

$$(23) \quad G(\mathbf{x}) = \begin{cases} \beta_{mn} |\mathbf{x}|^{2m-n} \log |\mathbf{x}| & \text{if } 2m - n \text{ is an even integer,} \\ \gamma_{mn} |\mathbf{x}|^{2m-n} & \text{otherwise,} \end{cases}$$

where β_{mn} and γ_{mn} are constants.

For example, for the thin plate spline, we have $m = n = 2$ so that $G(\mathbf{x}) = |\mathbf{x}|^2 \log |\mathbf{x}|$ and the polynomial p in (22) is of the first order (linear). Another familiar case is where $m = 2$ and $n = 3$. Then we have $G(\mathbf{x}) = |\mathbf{x}|$ and a linear polynomial for (22).

Regularization with RKHS norm Let us consider another regularization problem, where, instead of $J_m^n(f)$ in (12) and (13), we take $\sum_{m \geq 0} a_m J_m^n(f)$ with a_m being constants and $a_0 \neq 0$. According to the scenario used in establishing the above theorems, we have the following result in which the regularization problem is directly solved in an RKHS.

1. We can find a Green's function for the operator

$$\sum_{m \geq 0} (-1)^m \Delta^m.$$

The solution is then given by $f(\mathbf{x}) = \sum_{k=1}^N c_k G(\mathbf{x} - \mathbf{x}_k)$. Note that this time we do not need a polynomial term like (22).

2. *Gaussian RBF.* In a particular case, where $a_m = \frac{\sigma^{2m}}{m!2^m}$ ($\sigma > 0$), we have the Green's function as $G(\mathbf{x}) = c \exp(-\frac{\|\mathbf{x}\|^2}{2\sigma^2})$.

4. FROM RKHS TO GAUSSIAN PROCESSES

GPR (8) and (9) can be derived using the RKHS duality [18] using an appropriate choice of kernel. Here we derive (9) through the regularization problem (16)⁷.

Let $\{P, \Omega, \mathcal{B}\}$ be a probability space. Suppose then that a zero-mean Gaussian process $u(\mathbf{x}) \equiv u(\mathbf{x}, \omega)$ is given on the probability space, with $\mathbf{x} \in E$ and $\omega \in \Omega$. Consider the space \mathbf{Z} consisting of stochastic variables, which is induced by the Gaussian process:

$$\mathbf{Z} = \left\{ \sum_{j=1}^k a_j u(\mathbf{x}_j); a_j \in \mathbb{R}, \mathbf{x}_j \in E, k \in \mathbb{N} \right\}.$$

We then know that \mathbf{Z} is a pre-Hilbert space having the inner product

$$\langle Z_1, Z_2 \rangle = \int_{\Omega} Z_1(\omega) Z_2(\omega) dP(\omega) = \mathbf{E}[Z_1, Z_2]$$

where $Z_1, Z_2 \in \mathbf{Z}$. The completion of \mathbf{Z} is therefore a Hilbert space.

The covariance function $\Sigma_u(\mathbf{x}, \mathbf{y}) = \mathbf{E}[u(\mathbf{x})u(\mathbf{y})]$ induces a Kernel function on $E \times E$, so that Theorem 3 asserts that there exists RKHS \mathbf{H} associated with this kernel function. It's easy to see that the completion of \mathbf{Z} is isometrically isomorphic to \mathbf{H} [18].

Now again let us assume that the data $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$ are observed in the presence of an independent identically distributed observation noise. Consider the regularization problem (16) in the above \mathbf{H} . According to the *Representer Theorem*, the solution f is represented in the form (17). Considering the reproducing property (1), the norm $\|f\|_{\mathbf{H}}^2$ can therefore be written in the form

$$\|f\|_{\mathbf{H}}^2 = \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}$$

with α_i gathered into a column vector $\boldsymbol{\alpha}$ and $K(\mathbf{x}_i, \mathbf{x}_j)$ gathered into a matrix \mathbf{K} . We will rewrite the discrete regularization problem (16) as

$$\min_f \|\mathbf{y} - \mathbf{K} \boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}$$

with \mathbf{y} being a vector containing data values y_i .

Then expanding the square and taking the derivative with respect to $\boldsymbol{\alpha}$, we have

$$\frac{d}{d\boldsymbol{\alpha}} = 0 = -2\mathbf{K}^T \mathbf{y} + 2\mathbf{K}^T \mathbf{K} \boldsymbol{\alpha} + 2\lambda \mathbf{K} \boldsymbol{\alpha}.$$

Because \mathbf{K} is symmetric, it follows that

$$\begin{aligned} \mathbf{K} \mathbf{K} + \lambda \mathbf{K} \boldsymbol{\alpha} &= \mathbf{K} \mathbf{y} \\ \Leftrightarrow (\mathbf{K} + \lambda \mathbf{I}) \boldsymbol{\alpha} &= \mathbf{y} \\ \Leftrightarrow \boldsymbol{\alpha} &= (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}. \end{aligned}$$

Here we assume that \mathbf{K} and $\mathbf{K} + \lambda \mathbf{I}$ are invertible. Now substitute this expression for $\boldsymbol{\alpha}$ into $\mathbf{f} = \mathbf{K} \boldsymbol{\alpha}$,

$$\mathbf{f} = \mathbf{K} (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}.$$

This is an expression to synthesize the Gaussian process at all the locations corresponding to \mathbf{y} . For a particular data point $y_i = f(\mathbf{x}_i)$ this specializes to

$$y_i = [K(\mathbf{x}_i, \mathbf{x}_1), K(\mathbf{x}_i, \mathbf{x}_2), \dots, K(\mathbf{x}_i, \mathbf{x}_n)] (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}$$

Then generalizing from location \mathbf{x}_i to an arbitrary location \mathbf{x} and choosing $\mathbf{K} = \boldsymbol{\Sigma}^{-1}$ we have (9). It is interesting to note that the regularization parameter λ in RKHS functional regression takes on the role of the observation noise variance in (9). This is somewhat intuitive in that the functional regression requires increased regularization in the presence of increased noise.

⁷As for (8), see [18].

5. AN EQUIVALENCE BETWEEN RADIAL BASIS AND GAUSSIAN PROCESS MODELS

The RBF and GPR models initially seem quite different. GPR is a weighted sum of the data, whereas RBF is a weighted sum of a kernel indexed by the distances between the data. The RBF weights can be computed in advance, whereas the GPR weights seem to need to be computed at run time. However, since the problem of fitting functions in an RKHS to data yields both Radial Basis function and Gaussian process models, there should be a direct and simple relationship between the RBF and GPR models. In this section we state this relationship.

The solution of the system (6) for the RBF weights can be written

$$\mathbf{w} = \mathbf{G}^{-1}\mathbf{y}$$

with \mathbf{G} being the matrix containing $G(\mathbf{x}_i - \mathbf{x}_j)$ in the i, j th position. Substituting this expression for the weights, the RBF interpolation (5) can be written in vector-matrix notation as

$$\hat{y}_0 = \hat{f}(\mathbf{x}) = \mathbf{r}^T \mathbf{w} = \mathbf{r}^T \mathbf{G}^{-1} \mathbf{y}$$

with $\mathbf{r}^T = [G(\mathbf{x} - \mathbf{x}_1), G(\mathbf{x} - \mathbf{x}_2), G(\mathbf{x} - \mathbf{x}_3), \dots]$. If we take the data covariance as the RBF kernel this is exactly the same expression as (8). This equivalence is reasonable in some cases. In particular, the Gaussian distribution is both a prototypical choice of covariance and a standard choice for the RBF kernel.

With this equivalence in place, the main operational difference between the RBF and GPR techniques is the assumption on “what is known when”. In the RBF technique the data values y_i are assumed known in advance, whereas the locations to be interpolated (or extrapolated) are only known when the interpolation is evaluated. GPR has the opposite assumption, that the data covariance is known in advance (at arbitrary locations), but the surrounding data values y_i are not known until run-time.

On the other hand RBF and GPR models are not always equivalent. There are many choices of covariance that are not radially symmetric, and many choices of radial basis that cannot be covariances. A common instance of the latter was discussed in section 3.2, i.e., the radial basis functions generated as Green’s functions of a regularizer (13). These functions are zero at the origin, which precludes considering them as covariances.

6. CONCLUSION

RBF interpolation and Gaussian process regression can both be expressed in terms of a common underlying RKHS formulation. Specifically, both techniques are instances of functional regression in an RKHS. This common formulation illuminates the role and choice of the kernel, and allows us to identify an exact formal equivalence between the two techniques under appropriate choices of this kernel. While we wish to make this point of equivalence, the two techniques nevertheless have significant differences in

motivation and practice. The common formulation clarifies both similarities and differences.

ACKNOWLEDGMENTS

We thank Kenji Fukumizu, Geoffrey Irving and Jean-Marie Aubry for discussions of several points. This work is partially supported by Japan Science and Technology Agency, CREST Project.

REFERENCES

- [1] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.
- [2] W. Baxter and K. Anjyo, *Latent doodle space*, Computer Graphics Forum **25** (2006), 477–485.
- [3] S. Bergman, *The kernel function and the conformal mapping*, Mathematical Surveys **5** (1950).
- [4] D.S. Broomhead and D. Lowe, *Multivariable functional interpolation and adaptive networks*, Complex Systems **2** (1988), 321–355.
- [5] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans, *Reconstruction and representation of 3d objects with radial basis functions*, SIGGRAPH ’01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques (New York, NY, USA), ACM, 2001, pp. 67–76.
- [6] N. Cressie, *Statistics for spatial data*, Wiley, 1993.
- [7] Keith Grochow, Steven L. Martin, Aaron Hertzmann, and Zoran Popovic, *Style-based inverse kinematics*, ACM Trans. Graph. **23** (2004), no. 3, 522–531.
- [8] Tsuneya Kurihara and Natsuki Miyata, *Modeling deformable human hands from medical images*, Proceedings of the 2004 ACM SIGGRAPH Symposium on Computer Animation (SCA-04), 2004, pp. 357–366.
- [9] J. P. Lewis, *Generalized stochastic subdivision*, ACM Transactions on Graphics **6** (1987), no. 3, 167–190.
- [10] J. P. Lewis, Matt Cordner, and Nickson Fong, *Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation*, SIGGRAPH (New York, NY, USA), ACM Press/Addison-Wesley Publishing Co., 2000, pp. 165–172.
- [11] Christopher Lloyd, *Local models for spatial analysis*, CRC Press, 2007.
- [12] J. Meinguet, *Multivariate interpolation at arbitrary points made simple*, Z. Angew. Math. **30** (1979), 292–304.

- [13] Charles Micchelli, *Interpolation of scattered data: Distance matrices and conditionally positive definite functions*, Constructive Approximation (1986), no. 2, 11–22.
- [14] Tomohiko Mukai and Shigeru Kuriyama, *Geostatistical motion interpolation*, ACM Trans. Graph. **24** (2005), no. 3, 1062–1070.
- [15] Taehyun Rhee, J.P. Lewis, Ulrich Neumann, and Krishna Nayak, *Scan-based volume animation driven by locally adaptive articulated registrations*, IEEE Trans. Visualization and Computer Graphics **17** (2011), no. 3.
- [16] B. Schölkopf and A. J. Smola, *Learning with kernels*, MIT Press, 2002.
- [17] H. Todo, K. Anjyo, W. Baxter, and T. Igarashi, *Locally controllable stylized shading*, ACM Trans. Graph. **26** (2007), no. 3, 17.
- [18] Grace Wahba, *Spline models for observational data*, SIAM, 1990.
- [19] Holger Wendland, *Scattered data approximation*, Cambridge, 2005.
- [20] A. Yaglom, *An introduction to the theory of stationary random functions*, Dover, 1973.
- [21] K. Yosida, *Functional analysis (3rd ed.)*, Springer, 1971.

Ken Anjyo
OLM Digital, Inc.
Dens Kono Bldg. 302 ViPlus, Tokyo, Japan
E-mail: anjyo(at)acm.org

J. P. Lewis
Weta Digital and Victoria University, Wellington, New Zealand
E-mail: zilla(at)computer.org