ISSN 1995-4239, Numerical Analysis and Applications, 2018, Vol. 11, No. 3, pp. 220–235. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © A.I. Rozhenko, 2018, published in Sibirskii Zhurnal Vychislitel'noi Matematiki, 2018, Vol. 21, No. 3, pp. 273–290.

Comparison of Radial Basis Functions

A. I. Rozhenko*

Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch, Russian Academy of Sciences, pr. Akad. Lavrent'eva 6, Novosibirsk, 630090 Russia

Received November 8, 2017; in final form, February 10, 2018

Abstract—A survey of algorithms for approximation of multivariate functions with radial basis function (RBF) splines is presented. Algorithms of interpolating, smoothing, selecting the smoothing parameter, and regression with splines are described in detail. These algorithms are based on the feature of conditional positive definiteness of the spline radial basis function. Several families of radial basis functions generated by means of conditionally completely monotone functions are considered. Recommendations for the selection of the spline basis and preparation of initial data for approximation with the help of the RBF spline are given.

DOI: 10.1134/S1995423918030047

Keywords: *spline, algorithm, radial basis function, reproducing kernel, trend, external drift, interpolation, smoothing, regression, tension spline, regularized spline.*

INTRODUCTION

In this paper we discuss methods for approximating a multivariate function whose values are known in a finite set of measurement nodes. In other words, in some domain $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, a non-regular mesh of nodes is set in which the values of the function to approximate are known. The values can be measured with an error, and, thus, the approximating function is not expected to accurately reproduce (interpolate) the mesh data, i.e., methods of data smoothing are involved.

As a function approximation apparatus we will use splines constructed using some RBF. Under certain conditions (see below), the RBF is a reproducing kernel of some (semi-)Hilbert space, called a native space.[1] This fact provides a theoretical justification of convergence of RBF splines at condensation of meshes; having studied the properties of the RBF, one can draw conclusions about the smoothness of the functions of the native space and obtain estimates for the convergence of the RBF splines. Presence of native space also enables consideration of the smoothing problem and selection of the smoothing parameter with a natural criterion, when the quadratic deviation of the values of the smoothing spline at the mesh nodes from the measured values reaches a predetermined error threshold.

In this review, we confine ourselves to spline approximation algorithms the author has made some contribution to. Other approximation algorithms, as well as a comprehensive analysis of the theoretical aspects of approximation of multidimensional data, are given, for example, in [3].

1. RBF SPLINE

We will construct an approximation in the form of an RBF spline (see, for instance, [4, 5]):

$$\sigma(x) = \sum_{i=1}^{N} \lambda_i g(x, x_i) + p(x). \tag{1.1}$$

Here

*E-mail: rozhenko@oapmg.sscc.ru

[1] The term "native space" was apparently introduced in [1], and a way to build a native space was proposed in [2].

- $x_i \in \mathbb{R}^d$, $i = 1, \ldots, N$, are the *datum nodes* of the spline,
- $g(x, y) = \phi(|x - y|)$ is the *univariate radial basis function* ($\phi \in C[0, \infty)$),
- $|x - y|$ is the Euclidean distance between points $x, y \in \mathbb{R}^d$,
- $\lambda_i \in \mathbb{R}$ are the spline coefficients, and
- $p \in \mathcal{P}$, where $\mathcal{P}$ is a given finite-dimensional space of functions in $\mathbb{R}^d$, which is referred to as the *trend* of the spline. Usually, $\mathcal{P}$ is a space of algebraic polynomials of some degree.

In the interpolation and smoothing methods, the mesh of datum nodes coincides with that of measurement nodes, but in the general case these meshes are different. For example, in the spline regression method [6] the mesh of datum nodes is not related to the mesh of measurement nodes, whereas in the smoothing method with bilateral constraints [7, 8] the datum nodes of the spline are a submesh of the measurement nodes.

To obtain a closed system of linear equations we require additionally that the coefficients λ_i from (1.1) annihilate the trend of the spline:

$$\sum_{i=1}^{N} \lambda_i p(x_i) = 0 \qquad \forall\, p \in \mathcal{P}. \tag{1.2}$$

Then, having chosen some basis $u_i \in \mathcal{P}$, $i = 1, \ldots, K$, $K = \dim \mathcal{P}$, and writing down the trend function p from (1.1) as

$$p(x) = \sum_{i=1}^{K} \mu_i u_i(x),$$

under the interpolation conditions $\sigma(x_i) = z_i$, $i = 1, \ldots, N$, we obtain the following system of linear equations:

$$\begin{pmatrix} G & U \\ U^\top & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}. \tag{1.3}$$

Here G is an $N \times N$ matrix with coefficients $g_{ij} = g(x_i, x_j)$, U is an $N \times K$ matrix with coefficients $u_{ik} = u_k(x_i)$, $\lambda = (\lambda_1, \ldots, \lambda_N)^\top$ and $\mu = (\mu_1, \ldots, \mu_K)^\top$ are the vectors of the unknown spline coefficients, and $z = (z_1, \ldots, z_N)^\top$ is the vector of the values of the function at the measurement nodes.

We assume that all the nodes x_i of the measurement mesh are different, which is necessary for existence of a solution to problem (1.3).[2] For problem (1.3) to be uniquely soluble, the conditions $N \geq K$ and $\operatorname{rank} U = K$ must also be fulfilled. These conditions are also sufficient if the function $g(x, y)$ satisfies the conditions set out below.

[2] More precisely, if there are coinciding nodes, the measured values in them must also coincide. Otherwise, system (1.3) would not have solutions. However, in this situation coincident measurement nodes can simply be omitted.

2. CONDITIONAL POSITIVE DEFINITENESS

The function $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is referred to as *conditionally positively definite* in $\Omega \subset \mathbb{R}^d$ with respect to $\mathcal{P}$ if for any finite mesh of distinct points $y_i \in \Omega$, $i = 1, \dots, M$, $M \in \mathbb{N}$,

$$\sum_{i=1}^{M}\sum_{j=1}^{M} \xi_i \xi_j g(y_i, y_j) > 0, \tag{2.1}$$

where $\{\xi_i \in \mathbb{R},\ i = 1, \dots, M\}$ are all nontrivial sets of values satisfying the conditions

$$\sum_{i=1}^{M} \xi_i p(y_i) = 0 \quad \forall\, p \in \mathcal{P}. \tag{2.2}$$

The function $g(x, y)$ is not necessarily symmetric with respect to x and y. For example, with $d = 1$ the function $g(x, y) = (-1)^m (x - y)_+^{2m-1}$ is conditionally positively definite with respect to polynomials of degree $m - 1$, but it is not symmetric. Here and below $(t)_+ = \max(t, 0)$.

In the absence of symmetry, we can talk about *conditional symmetry* of the function g in Ω with respect to $\mathcal{P}$:

$$\sum_{i=1}^{M}\sum_{j=1}^{M} \xi_i \xi_j g(y_i, y_j) = \sum_{i=1}^{M}\sum_{j=1}^{M} \xi_i \xi_j g(y_j, y_i) \tag{2.3}$$

under conditions (2.2).

In the case of conditional positive definiteness and conditional symmetry, it can be asserted that the function $g(x, y)$ is a *reproducing kernel* of some semi-Hilbert space [9, 10], which is called a native space for the triple $\langle g, \Omega, \mathcal{P}\rangle$ and denoted as $\mathcal{N}_\phi(\Omega, \mathcal{P})$ when $g(x, y) = \phi(|x - y|)$.[3]

The idea of constructing a native space is very simple. In the subspace of all splines of form (1.1), (1.2) that satisfy the condition $p(x) \equiv 0$ and are constructed on all sorts of finite meshes of datum nodes from Ω, a bilinear form

$$(\sigma_1, \sigma_2)_G = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \lambda_i^{(1)} \lambda_j^{(2)} g(x_i^{(1)}, x_j^{(2)})$$

will be positively definite and symmetric by virtue of conditions (2.1)–(2.3). Performing completion of the linear space of such splines in the norm $\|\sigma\|_G = (\sigma, \sigma)_G^{1/2}$, we obtain a Hilbert space, which we denote by $G_g(\Omega, \mathcal{P})$. Then $\mathcal{N}_g(\Omega, \mathcal{P}) = G_g(\Omega, \mathcal{P}) + \mathcal{P}$ is the required native half-Hilbert space, and $\|\cdot\|_G$ is a seminorm in it with the kernel $\mathcal{P}$.[4]

It follows that the matrix G from (1.3) will be symmetric and positively definite on the subspace of vectors λ satisfying the condition $U^\top \lambda = 0$. This idea forms the basis of the algorithm for spline construction via converting system of equations (1.3) into a system with a symmetric positively definite matrix, also known as the Anselone–Laurent algorithm [11, 13].

Since g is a reproducing kernel of the function space $\mathcal{N}_g(\Omega, \mathcal{P})$ in the seminorm $\|\cdot\|_G$, spline (1.1), (1.2) with coefficients satisfying system of equations (1.3) minimizes the seminorm $\|\cdot\|_G$ on the set of all functions $f \in \mathcal{N}_g(\Omega, \mathcal{P})$ that satisfy the interpolation conditions $f(x_i) = z_i$, $i = 1, \dots, N$ (see, for instance, [9, 10]).

[3] In [3] and other sources, a simplified notation $\mathcal{N}_\phi(\Omega)$ is used, in which the second parameter is omitted.

[4] If we represent this seminorm as $\|f\|_G = \|Tf\|_Y$, where T is a linear operator acting into some Hilbert space Y (T can be, for example, a quotient map from $\mathcal{N}_g(\Omega, \mathcal{P})$ into $G_g(\Omega, \mathcal{P})$ with a kernel of $\mathcal{P}$), then $g(x, y)$ will be the Green function of the operator T^*T (see, for instance, [11, 12]).

Due to the minimization condition for the seminorm $\|\cdot\|_G$ on splines, we can also formulate the problem of constructing the smoothing spline:

$$\sigma_\alpha = \arg \min_{f \in \mathcal{N}_g(\Omega,\mathcal{P})} \left(\alpha \|f\|_G^2 + \sum_{i=1}^N \frac{(f(x_i) - z_i)^2}{p_i} \right). \tag{2.4}$$

Here $p_i > 0$ are some specified weight parameters proportional to the square of the measurement error for z_i values at the mesh nodes, and $\alpha > 0$ is the smoothing parameter, which can be varied. If we assume that system (1.3) has a single solution ($N \geq K$ and $\mathrm{rank} U = K$), then problem (2.4) has a unique solution, the coefficients of which satisfy the following system of equations:

$$\begin{pmatrix} G + \alpha P & U \\ U^\top & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}. \tag{2.5}$$

Here $P = \mathrm{diag}(p_1, \ldots, p_N)$ is the diagonal matrix of weights from (2.4). It is obvious that when $\alpha \to 0$, the smoothing spline σ_α tends to the interpolation one, and when $\alpha \to \infty$, in the limit we have the function from the trend $\mathcal{P}$ that minimizes the functional $\sum\limits_{i=1}^N \frac{(f(x_i)-z_i)^2}{p_i}$. This brings up the question how we should choose the smoothing parameter α?

Let

$$\rho(f) = \left(\sum_{i=1}^N \frac{(f(x_i) - z_i)^2}{p_i} \right)^{1/2} \tag{2.6}$$

denote the weighted quadratic deviation of the function $f \in \mathcal{N}_g(\Omega, \mathcal{P})$ from measurements at the mesh nodes. We will select the parameter α basing on the *residual principle*, solving the nonlinear equation $\rho(\sigma_\alpha) = \varepsilon$, where $\varepsilon > 0$ is the required weighted error threshold.

An algorithm for solving such an equation can be reduced to an iterative process, at each step of which problem (2.5) is solved with $\alpha = \alpha_n$, $n = 0, 1, \ldots$. Here α_0 is the initial approximation, and α_1, α_2, ... are approximations at subsequent steps of the iteration process. An algorithm for solving this problem, which can be reduced to iterations of the Newton method of the second order of convergence, was proposed in [14]; the optimal iteration method was justified in [15]; a combined algorithm using iterations of the Newton method and a fractional-rational approximation was proposed in [16, 17]. Finally, an algorithm with a higher (predetermined) degree of convergence was justified in [18] and brought to practical implementation in [19].

3. SELECTING THE SMOOTHING PARAMETER

We will describe the algorithm for choosing the smoothing parameter from the residual principle in accordance with [17, 19]. We introduce the residual operator R_α, which relates the vector of measurements z to the residual vector $\zeta = R_\alpha z$ by the formula $\zeta_i = z_i - \sigma_\alpha(x_i)$, $i = 1, \ldots, N$, where σ_α is smoothing problem solution (2.4). We also introduce a weighted scalar product in $Z = \mathbb{R}^N$: $(u, v)_Z = \sum\limits_{i=1}^N \frac{u_i v_i}{p_i}$. Then the residual equation can be represented as follows:

$$\rho(\sigma_\alpha) = \|R_\alpha z\|_Z = \varepsilon. \tag{3.1}$$

From the first group of equations in (2.5) we derive that $G\lambda + U\mu = z - \alpha P\lambda$. The left-hand side is the vector of the values of the smoothing spline at the mesh nodes. Therefore, $R_\alpha z = \alpha P \lambda$.

It is obvious that the function $\rho(\sigma_\alpha)$ increases monotonically, the monotonicity being strict if $\lambda \neq 0$, i.e., if the equation $U\mu = z$ has no solutions (the measurement vector cannot be interpolated by a

function from the trend). In other words, either $\rho(\sigma_\alpha)$ is a strictly monotonically increasing function or $\rho(\sigma_\alpha) \equiv 0$. In what follows we assume that $\rho(\sigma_\alpha) \not\equiv 0$.

It is also obvious that Eq. (3.1) has a solution if $\varepsilon \le \varepsilon_{\max} = \|R_\infty z\|_Z$. We can obtain the $\varepsilon_{\max}$ value by solving the problem of weighted least squares on $\mathcal{P}$ or passing to the limit with $\alpha \to \infty$ in system of equations (2.5) as follows: setting $\beta = 1/\alpha$ and $\tilde{\lambda} = \alpha\lambda$ and multiplying the second group of equations by α, we obtain an equivalent system of equations:

$$\begin{pmatrix} \beta G + P & U \\ U^\top & 0 \end{pmatrix} \begin{pmatrix} \tilde{\lambda} \\ \mu \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}. \tag{3.2}$$

Therefore, $R_\infty z = P\tilde{\lambda}$ at solution to (3.2) with $\beta = 0$.

Let $0 < \varepsilon < \varepsilon_{\max}$, α_* be a solution to Eq. (3.1), and α_0 be the initial approximation of the smoothing parameter. According to [15], Eq. (3.1) should be converted to the form $1/\rho(\sigma_{1/\beta}) = 1/\varepsilon$, with which the rate of convergence of the Newton method is maximal. The function $1/\rho(\sigma_{1/\beta})$ is strictly convex upwards, and thus the iterations of the Newton method will converge if $\alpha_0 \le \alpha_*$, i.e., $\|R_{\alpha_0} z\|_Z \ge \varepsilon$. If this condition is not met, then, as the first step, we construct a fractional-rational approximation $\tilde{\rho}(\alpha)$ of the residual function from the three values: $\rho(\sigma_{\alpha_0})$, $\rho'(\sigma_{\alpha_0})$, and $\rho(\sigma_\infty) = \varepsilon_{\max}$, and choose α_1 by solving the residual equation $\tilde{\rho}(\alpha) = \varepsilon$.

The formula for the step of the Newton method for $\rho(\sigma_{\alpha_n}) \ge \varepsilon$ is as follows:

$$\alpha_{n+1} = \alpha_n \frac{1 - \omega(\alpha_n)}{\rho(\sigma_{\alpha_n})/\varepsilon - \omega(\alpha_n)}, \tag{3.3}$$

$$\omega(\alpha) = \frac{(R_\alpha z, R_\alpha^2 z)_Z}{(R_\alpha z, R_\alpha z)_Z} = \frac{(R_\alpha z, R_\alpha^2 z)_Z}{\rho^2(\sigma_\alpha)}.$$

The formula for the step of the fractional-rational method with $\rho(\sigma_{\alpha_n}) < \varepsilon$ is as follows:

$$\alpha_{n+1} = \alpha_n \frac{(1 - \rho(\sigma_{\alpha_n})/\varepsilon_{\max}) - (1 - \rho(\sigma_{\alpha_n})/\varepsilon)\, d(\alpha_k)}{\rho(\sigma_{\alpha_n})/\varepsilon - \rho(\sigma_{\alpha_n})/\varepsilon_{\max}}, \tag{3.4}$$

$$d(\alpha) = \frac{\rho(\sigma_\alpha)/\varepsilon_{\max} - \omega(\alpha)}{1 - \omega(\alpha)}.$$

At each step of the iterations by formulas (3.3) and (3.4), it is required to solve two systems of linear equations of form (2.5) with the vectors z and $R_\alpha z$ in the right-hand side. When direct methods of solving system of equations (2.5) are used, the time complexity of factorization of the matrix is much higher than the time complexity of solving the system with a factorized matrix. Therefore, we can win in the number of operations if we achieve a higher rate of convergence of the method via increasing the number of problems of form (2.5) solved at each iteration.

Such method is based on the expansion of the operator R_α and its complementary operator Q_β [17, 18]:

$$R_\alpha = I - \sum_{m=0}^{\infty} \Big(\frac{\alpha_0 - \alpha}{\alpha_0}\Big)^m R_{\alpha_0}^m (I - R_{\alpha_0}) = \frac{\alpha}{\alpha_0} \sum_{m=0}^{\infty} \Big(\frac{\alpha_0 - \alpha}{\alpha_0}\Big)^m R_{\alpha_0}^{m+1}, \tag{3.5}$$

$$R_{1/\beta} = \sum_{m=0}^{\infty} \Big(\frac{\beta_0 - \beta}{\beta_0}\Big)^m Q_{\beta_0}^m (I - Q_{\beta_0}) = I - \frac{\beta}{\beta_0} \sum_{m=0}^{\infty} \Big(\frac{\beta_0 - \beta}{\beta_0}\Big)^m Q_{\beta_0}^{m+1}. \tag{3.6}$$

Replacing in these formulas the upper summation limit with $L-1$ at some finite L, we obtain two-sided approximations with the order L of the residual function in neighborhood of the points α_0 and β_0 [18], and hence the algorithm using these approximations will have the order of convergence L. A detailed study of ways of approximation by this method was performed in [19]. The results are presented below.

If $\rho(\sigma_{\alpha_n}) \geq \varepsilon$, using the formula

$$R_\alpha^{(L)} = \frac{\alpha}{\alpha_n} \sum_{m=0}^{L-1} \Big(\frac{\alpha_n - \alpha}{\alpha_n}\Big)^m R_{\alpha_n}^{m+1}$$

and solving the equation $\|R_\alpha^{(L)} z\|_Z = \varepsilon$, we obtain α_{n+1}.

If $\rho(\sigma_{\alpha_n}) < \varepsilon$, at the first step ($n=0$) we select α_1 by fractional-rational approximation method (3.4), and with other n we calculate $\alpha_{n+1} = 1/\beta_{n+1}$ by solving the equation $\|\hat{R}_{1/\beta}^{(L)} z\|_Z = \varepsilon$, where

$$\hat{R}_{1/\beta}^{(L)} = \sum_{m=0}^{L-1} \Big(\frac{\beta_n - \beta}{\beta_n}\Big)^m Q_{\beta_n}^m (I - Q_{\beta_n}).$$

It is easy to see that in both cases at each step it is required to solve L systems like (2.5) with different right-hand sides. It is clear that when calculating $\|R_\alpha^{(L)} z\|_Z$ ($\|\hat{R}_{1/\beta}^{(L)} z\|_Z$) we can also calculate $\|R_\alpha^{(l)} z\|_Z$ ($\|\hat{R}_{1/\beta}^{(l)} z\|_Z$) for $l < L$. Based on this information, we can extrapolate approximations of the residual function in L and refine its value for each α (β), i.e., accelerate the convergence. This method was applied in [20] and then improved in [19].

4. ALGORITHM FOR SOLVING THE SMOOTHING PROBLEM

It is clear that the system of equations in interpolation problem (1.3) is a particular case of system of smoothing equations (2.5). Therefore, it suffices to consider the general case. System (2.5) is not symmetric if the function $g(x,y)$ is not symmetric (the symmetry is obvious in the case of RBF splines). Even if the matrix G is symmetric, the spectrum of this system is alternating (of course, with a nontrivial trend $\mathcal{P}$). Another shortcoming of this system is that the conditioning number of its matrix depends on the choice of basis in $\mathcal{P}$. Therefore, application of standard solution methods with a poor choice of basis in $\mathcal{P}$ can make the solution less accurate by far.

Under the assumptions of Section 2, the matrix G is symmetric and positively definite on the subspace of vectors such that $U^\top \lambda = 0$. Therefore, we would like to take advantage of this fact and reduce system (2.5) to an equivalent system of equations on the subspace with a symmetric and positively definite matrix. Such a transformation was actually proposed in [13]: it was based on building a linear operator $H : \mathbb{R}^N \to \mathbb{R}^{N-K}$ with $\operatorname{rank} H = N-K$ and $HU = 0$, i.e., this operator must be of full rank and annihilate the image of the operator U. In the classical case of univariate polynomial splines of degree $2m-1$, the matrix of the operator H was collected from the coefficients of finite differences of the mth order that annihilate polynomials of degree $m-1$ on different submeshes. In the multivariate case, many researchers used a similar approach, but the result was strongly dependent on the choice of submeshes for constructing finite differences and did not guarantee that the conditioning of the initial system would not worsen under such a transformation.

The author suggested in [21] an algorithm of purely algebraic construction of the operator H that preserved the conditioning number of the matrix G ($\operatorname{cond}_2 G$) on the subspace of vectors such that $U^\top \lambda = 0$. We perform upper triangularization of the matrix U by the method of rotations or reflections:

$$QU = \begin{pmatrix} \tilde{U} \\ O \end{pmatrix}. \tag{4.1}$$

Here Q is the transformation operator, consisting of a sequence of elementary operations of rotations or reflections, $\tilde{U}$ is a $K \times K$ nondegenerate upper triangular matrix, and O is a zero block of dimension $(N-K) \times K$. It is clear that the image of the operator QU consists of vectors the nonzero elements of which are in the first K positions, and to annihilate the image we should just omit the first K components of the vector. Formally, we split the matrix of the identity operator I in $\mathbb{R}^N$ into two blocks:

$$I = \begin{pmatrix} J_1 \\ J_2 \end{pmatrix},$$

where $J_1 : \mathbb{R}^N \to \mathbb{R}^K$ is composed of the first K rows of the matrix of the operator I, and $J_2 : \mathbb{R}^N \to \mathbb{R}^{N-K}$ of the rest $N-K$ rows. Then the sought-for operator H is $H = J_2 Q$.

From the identity $U^\top \lambda = 0$ it follows that the vector λ belongs to the kernel of the operator $U^\top$, and thus it is easy to represent the vector λ as $H^\top \nu$, the vector $\nu \in \mathbb{R}^{N-K}$ determined uniquely. Substituting $\lambda = H^\top \nu$ into (2.5) and applying the operator H to the first group of equations, we obtain an equivalent system of equations with respect to ν:

$$(A + \alpha H P H^\top)\nu = Hz, \tag{4.2}$$

where $A = HGH^\top$. It is clear that $(A\nu, \nu) = (G\lambda, \lambda)$ and $U^\top \lambda = 0$ by virtue of the properties of the operator H. So, we conclude that A is symmetric and positively definite. Its conditioning number is the same as that of the matrix G on the subspace $U^\top \lambda = 0$ because orthogonal transformations are applied.

Having solved this system of equations, we find λ from the identity $\lambda = H^\top \nu$ as follows: We add K zeros to the beginning of the vector ν, i.e., apply the transformation $J_2^\top$, and then we execute the sequence of transposed elementary operations composing the transformation Q, but in the reverse order. We easily found the vector μ from the identity $U\mu = z - (G + \alpha P)\lambda$ via pre-multiplying it by $J_1 Q$ and substituting $\lambda = H^\top \nu$:

$$\tilde{U}\nu = J_1 Q z - J_1 Q (G + \alpha P) H^\top \nu. \tag{4.3}$$

The matrix of this system is upper triangular, and the vector $J_1 Q z$ and matrix $J_1 Q (G + \alpha P) H^\top$ are found "for free" in the course of the transformation of system (2.5) into (4.2).

Remark. The substitution algorithm can also be applied to solving a general system of equations with a saddle point of the following form:

$$\begin{pmatrix} B & U \\ U^\top & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} z \\ w \end{pmatrix}, \tag{4.4}$$

where B is an $N \times N$ matrix and U is a full-rank $N \times K$ matrix, $K \le N$. We again use the orthogonal transformation Q to reduce the matrix U to upper triangular form (4.1) and perform the substitution $\lambda = Q^\top \nu$ to split the vector ν into two parts: ν_1, which includes the first K components of the vector ν, and ν_2, which includes the rest $(N-K)$ components. Pre-multiplying the first group of equations in (4.4) by Q and decomposing $QBQ^\top$ into four blocks C_{ij}, $i, j = 1, 2$, with the diagonal blocks C_{11} and C_{22} of the dimensions $K \times K$ and $(N-K) \times (N-K)$, respectively, we arrive at the following system of equations:

$$\begin{pmatrix} C_{11} & C_{12} & \tilde{U} \\ C_{21} & C_{22} & 0 \\ \tilde{U}^\top & 0 & 0 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \mu \end{pmatrix} = \begin{pmatrix} J_1 Q z \\ J_2 Q z \\ w \end{pmatrix}.$$

From the last group of equations with the lower triangular matrix $\tilde{U}^\top$ we find ν_1; then from the second group of equations with the square matrix C_{22} we find ν_2; further, from the first group of equations with the upper triangular matrix $\tilde{U}$ we determine μ; finally, we calculate $\lambda = Q^\top \nu$.

5. SPLINE REGRESSION ALGORITHM

Now we consider the case when the mesh of datum nodes of spline differs from the mesh in which the values of the function to approximate are known. In the spline regression method, we will construct a spline of form (1.1), (1.2), basing on the values z_i given on the mesh y_i, $i = 1, \ldots, M$, $M > N$. It is clear that the system of interpolation equations and Eq. (1.2) is most likely to be disjoint, and only an approximate solution that minimizes the weighted residual functional $\sum\limits_{i=1}^{M} \frac{(\sigma(y_i)-z_i)^2}{p_i}$ can be obtained. In terms of matrices, we need to deduce a pseudo-solution to the following system of equations:

$$D(F\lambda + V\mu) = Dz, \tag{5.1}$$

where $D = \operatorname{diag}(p_1^{-1/2}, \ldots, p_M^{-1/2})$ is the diagonal matrix of weighting coefficients, F is the $M \times N$ matrix with the coefficients $f_{ij} = g(y_i, x_j)$, and V is the $M \times K$ matrix with the coefficients $v_{ik} = u_k(y_i)$. We will additionally require strict fulfillment of conditions (1.2), i.e., the vector λ is to satisfy the linear constraints $U^\top \lambda = 0$. As a result, we arrive at the linear-equality-constraint least-squares problem.

Performing the substitution $\lambda = H^\top \nu$, we get rid of the linear constraints and obtain the following system of equations:

$$D(FH^\top \nu + V\mu) = Dz,$$

the solution of which is found by the usual method of least squares.

The spline regression method is useful when the number of points with measurements is too large for construction of a smoothing spline on the entire mesh of measurements or if the mesh of measurements is very uneven or the measurements are contradictory. This brings up the question how we should choose the datum nodes of the spline.

One way is to thin out the measurement mesh. We choose some radius of thinning $r_0 > 0$ and require absence of other nodes from the measurement mesh in the r_0 neighborhood of each node of the thinned-out mesh. A decision on selection of the best node in the r_0 neighborhood will be taken from the measurement quality criterion calculated at each node. The quality criterion will be the probability of the event that the measured value of the function in a node is a data outlier. The probability is evaluated using the Grubbs statistical test [22].

Let ξ_i, $i = 1, \ldots, n$, be a sample of n values of normally distributed random value, and $\tilde{\xi}$ be one of the values in this sample. It is required to estimate the probability $\tau(\tilde{\xi})$ of the event that this value is a data outlier. We introduce $Z = (\tilde{\xi} - \bar{\xi})/s$, where $\bar{\xi}$ and s are the average value and standard deviation obtained on this sampling, respectively. Then Z will be a data outlier with a significance level $\alpha > 0$ in one-sided Grubbs test if

$$Z^2 > \frac{(n-1)^2}{n} \frac{t^2_{\alpha/n,n-2}}{n-2+t^2_{\alpha/n,n-2}},$$

where $t_{\alpha/n,n-2}$ is a solution to the equation $t_{n-2}(x) = 1 - \alpha/n$ and $t_{n-2}(x)$ is the Student distribution function with $(n-2)$ degrees of freedom. In this formula, it is the calculation of the significance level α with respect to Z and n that is of interest to us.

With notation $Y = \frac{Z^2 n}{(n-1)^2}$, we obtain that

$$1-\frac{\alpha}{n}<t_{n-2}\left(\sqrt{\frac{Y(n-2)}{1-Y}}\right).$$

From this we deduce that the probability of data outlier is estimated as $1-\alpha$, i.e., we set that

$$\tau(\tilde{\xi})=\max\left\{1-n\left(1-t_{n-2}\left(\sqrt{\frac{Y(n-2)}{1-Y}}\right)\right),0\right\}.$$

In this formula it is assumed that $n>2$ (for $n=2$ we set that $\tau(\tilde{\xi})=0$). It is easy to verify that $Y=1$ in an extreme case, when other ξ_i except for $\tilde{\xi}$ are equal one to another. Then $\tau(\tilde{\xi})=1$. Otherwise, $0\le Y<1$ and $\tau(\tilde{\xi})<1$.

Formally, the Grubbs test cannot be applied to estimating the probability of outlier in data obtained on a mesh of nodes in $\mathbb{R}^d$, but in this test, the exact value of the outlier probability is of no interest to us. We want to compare the probabilities for different nodes in order to choose the best of them.

The Grubbs test can be applied to measurements the nodes of which are in the r_1 neighborhood of each node y_i, and from this sample we calculate the probability of outlier in the node y_i. We choose a rather large estimation radius $r_1>r_0$ and vary the radius r_0, which allows us to thin out the mesh with different levels of detailing. Experiments the author carried out to approximate the universal characteristic of hydroturbine (unpublished so far) allowed us to conclude that the behavior of a spline constructed by the spline regression method with a thinned-out mesh following this algorithm is rather close to the behavior of a smoothing spline with a similar level of weighted quadratic deviation, the smoothing results being still somewhat better than the spline-regression ones.

Note that a spline constructed by the spline regression method contains fewer coefficients than a smoothing spline. Therefore, such a solution is more reliable and its predictive ability is higher than that of a smoothing spline.[5]

6. SPLINE APPROXIMATION WITH EXTERNAL DRIFT

Let us extend the vector space of the trend $\mathcal{P}$ by supplementing it with some linearly independent basis functions. We denote the extended vector space with $\mathcal{Q}$. It is clear that if the function $g(x,y)$ is conditionally positively definite and symmetric with respect to $\mathcal{P}$, then it is also conditionally positively definite and symmetric with respect to $\mathcal{Q}$. That is, $g(x,y)$ will be a reproducing kernel of the space $\mathcal{N}_g(\Omega,\mathcal{Q})$, and all kinds of spline with the trend $\mathcal{Q}$ can be constructed. It is also obvious that $\mathcal{N}_g(\Omega,\mathcal{P})=G_g(\Omega,\mathcal{P})+\mathcal{P}$ and $\mathcal{N}_g(\Omega,\mathcal{Q})=G_g(\Omega,\mathcal{Q})+\mathcal{Q}$ are in the following relation:

$$G_g(\Omega,\mathcal{P})\supset G_g(\Omega,\mathcal{Q}),\qquad \mathcal{P}\subset\mathcal{Q},$$

i.e., the spline part of the native space narrows as the trend is extended.

Usually $\mathcal{P}$ consists of polynomials of a certain degree, and quite arbitrary functions can be added in its extension $\mathcal{Q}$. For example, if it is possible to construct a rough estimate of function to approximate using some methods, a rough estimation function or a set of such functions can be added to the basis of the trend. Functions that describe a data gap can also be added, i.e., we get an opportunity to approximate a function with discontinuities using RBF splines with a special extension of the trend. Such splines are called *splines with external drift* by analogy with the dual Kriging method with external drift (see, for instance, [24]).

[5] In the methods of machine learning they try to use approximation with a minimum number of variable coefficients. Of course, increasing the complexity of the approximator (the number of variable coefficients), one can better approximate the function values on a training sample, but this can be tied with "the overfitting" effect, when the approximation accuracy on a test sample deteriorates much. Therefore, if two approximators give results of similar accuracy, it is preferable to use the approximator of less complexity. On the other hand, if there are several fundamentally different approximators, their combination can result in a more reliable approximator. Namely this approach was used by the team that won the prize of the Netflix company [23]. Two dominating teams combined their algorithms and achieved the necessary prediction accuracy.

7. CONDITIONAL COMPLETE MONOTONICITY

In what follows we assume that the boundary of the domain Ω of the function to approximate is not essential for the approximation, an thus we set that $\Omega = \mathbb{R}^d$. Let P_n^d denote a vector space of multivariate algebraic polynomials in $\mathbb{R}^d$ of degree n. We set that $P_{-1}^d = \{0\}$. At $d = 1$, we write simply P_n.

Provided that conditions (2.1) are satisfied and $\mathcal{P} = P_{m-1}^d$, $m \in \mathbb{Z}_+$, a function $g(x, y)$ is called *conditionally positively definite of order* m. From Section 6 it follows that conditional positive definiteness will also hold for any integer $n > m$.

A function $\phi \in C[0, \infty)$ is also called *conditionally positively definite in* $\mathbb{R}^d$ if its respective radial basis function $g(x, y) = \phi(|x - y|)$ is conditionally positively definite in $\mathbb{R}^d$. The symmetry of radial basis function is obvious.

Let $\mathcal{R}_m^d$ denote the set of all functions $\phi \in C[0, \infty)$ that are conditionally positively definite of order m in $\mathbb{R}^d$. It is clear that $\mathcal{R}_{m_1}^{d_1} \subset \mathcal{R}_m^d$ at $m_1 \le m$, $d_1 \ge d$. We introduce the following notation:

$$\mathcal{R}_m^\infty = \bigcap_{d=1}^{\infty} \mathcal{R}_m^d.$$

The nonemptiness of the set $\mathcal{R}_m^\infty$ will follow from the reasoning below.

Functions from $\mathcal{R}_m^\infty$ can be used for construction of splines with any number of variables using a polynomial trend of the same degree $m - 1$. For example, with $m = 2$ the trend will consist of linear functions, and the dimensionality of the space P_1^d depends linearly on d. Thus, using such splines we can avoid "the curse of dimensionality" when the complexity of approximating function grows exponentially with d.

The support of functions from $\mathcal{R}_m^\infty$ is nonlocal [3, Corollary 9.3], and the spline value by formula (1.1) has to be calculated via summation over all datum nodes. With a finite d, in $\mathcal{R}_m^d$ there are functions with local support, an effective algorithm for constructing which is described, for example, in [3].

A function $f : \mathbb{R}_+ \to \mathbb{R}$ is called *completely monotone*[6] *of order* m if $f \in C^\infty(0, \infty)$ and $(-1)^k f^{(k)} \ge 0$ for all $k \ge m$. We denote the set of such functions by $\mathcal{M}_m$, $m \in \mathbb{Z}_+$. The sets $\mathcal{R}_m^\infty$ and $\mathcal{M}_m$ are closely tied:

- $f \in \mathcal{M}_m \setminus P_m$ and $|f(0)| < \infty \implies f(r^2) \in \mathcal{R}_m^\infty$;
- $\phi \in \mathcal{R}_m^\infty \implies \phi(\sqrt{\cdot}) \in \mathcal{M}_m \setminus P_m$.

Sufficient conditions for belonging to the set $\mathcal{R}_m^\infty$, as well as necessary conditions at $m = 0$, were obtained in [25]. Later the necessary conditions were proved for all m (see details in [3]).

If $f \in \mathcal{M}_m$, then $bf(a^2(t + c^2)) + p_{m-1}(t)$ also belongs to $\mathcal{M}_m$ at $a \ne 0$, $b > 0$, $p_{m-1} \in P_{m-1}$. This results in the following generation rule for a function $\phi \in \mathcal{R}_m^\infty$:

$$f \in \mathcal{M}_m \setminus P_m \implies \phi(r) = bf(a^2(r^2 + c^2)) + p_{m-1}(r^2) \implies \phi \in \mathcal{R}_m^\infty, \tag{7.1}$$

and $c \ne 0$ if the function f is unbounded at zero.

In approximation with RBF splines, the following functions are often used in the construction of radial basis functions:

- multiquadric: $\phi(r) = (-1)^{\lfloor \nu \rfloor + 1}(r^2 + c^2)^\nu$, $\nu \in \mathbb{R}_+ \setminus \mathbb{Z}_+$;
- inverse multiquadric: $\phi(r) = (r^2 + c^2)^\nu$, $\nu < 0$;

[6] We use the terms "completely monotone" and "complete monotonicity" as synonyms of the terms "completely continuous" and "complete continuity".

- polyharmonic RBF [26, 27]: $\phi(r) = (-1)^{n+1} r^{2n} \ln r$, $n \in \mathbb{N}$;
- power RBF [26, 27]: $\phi(r) = (-1)^{\lfloor \nu \rfloor + 1} r^{2\nu}$, $\nu \in \mathbb{R}_+ \setminus \mathbb{Z}_+$;
- RBF of DMM spline [28, 29]: $\phi(r) = (-1)^{n+1}(r^2 + c^2)^n \ln(r^2 + c^2)$, $n \in \mathbb{Z}_+$.

All these RBFs are generated with the help of the following family of conditionally completely monotonic functions:

$$f_\nu(t) = \begin{cases} \Gamma(-\nu) t^\nu, & \nu \in \mathbb{R} \setminus \mathbb{Z}_+, \\ (-1)^{\nu+1} t^\nu \ln t, & \nu \in \mathbb{Z}_+, \end{cases} \qquad f_\nu \in \mathcal{M}_{(\lfloor \nu \rfloor + 1)_+}. \tag{7.2}$$

Here and below $\Gamma(a)$ is a gamma function. Recall that $(k)_+ = \max(k, 0)$. At $\nu \le 0$ the function f_ν is unbounded at zero, and, thus, the Hardy parameter c in the generation of RBF must be non-nil in this case.

8. COMBINING RADIAL BASIS FUNCTIONS

Let $\phi_i \in \mathcal{R}^{d_i}_{m_i}$, $i = 1, \dots, n$, and $\phi = \sum\limits_{i=1}^{n} a_i \phi_i$ at $a_i > 0$. Then $\phi \in \mathcal{R}^d_m$, where $m = \max_i m_i$, $d = \min_i d_i$. Some linear combinations of the RBFs with nonpositive coefficients also lead to conditionally positively definite RBFs. For example, radial basis functions of tension splines and regularized splines [30, 31] can be obtained via subtraction of several RBFs.

A proof of conditional positive definiteness of radial basis functions of certain tension splines and regularized splines from [30] was presented in [32] and generalized in [33]; the algorithm for obtaining radial basis functions of completely regularized splines from [33] was generalized in [34].

The results of the generalization are given below in accordance with [34] and are stated in terms of generators of conditionally completely continuous functions.

Kernels of Sobolev's Spaces and Their Generalization

Let us consider a family of functions

$$h_\nu(t) = t^{\nu/2} K_\nu(\sqrt{t}), \quad \nu \in \mathbb{R}, \quad t \in (0, \infty).$$

Here $K_\nu(x)$ is a modified Bessel function of the second kind of order ν [35]. The function h_ν generates a reproducing kernel of Sobolev's space $H^k(\mathbb{R}^d)$ at $\nu = k - d/2$, which is known as the Whittle–Matérn RBF [36].

Here the generalization consists in extending the representation h_ν to any ν, including negative ones. The functions h_ν belong to $\mathcal{M}_0$ for all $\nu \in \mathbb{R}$ and are bounded at zero when $\nu > 0$.

Generalization of Tension Splines and Regularized Splines

Let us consider an auxiliary family of functions

$$\tilde{h}_\nu(t) = \begin{cases} \dfrac{\Gamma(-\nu) t^\nu}{2^{\nu+1}}, & \nu \in \mathbb{R} \setminus \mathbb{Z}_+, \\ (-1)^{\nu+1} \dfrac{t^\nu [\ln(t/4) - \psi(1) - \psi(\nu+1)]}{\nu!\, 2^{\nu+1}}, & \nu \in \mathbb{Z}_+, \end{cases}$$

where $\psi(n) = -\gamma + \sum\limits_{k=1}^{n-1} k^{-1}$ is the logarithmic derivative of the gamma function, and γ is the Euler constant. Comparing $\tilde{h}_\nu$ with f_ν from (7.2), we can note easily that they differ by a positive factor, and

when $\nu \in \mathbb{Z}_+$, a constant is also added, which does not affect the properties of the conditional complete monotonicity. It is clear that $\tilde{h}_\nu \in \mathcal{M}_{(\lfloor\nu\rfloor+1)_+}$ and $\tilde{h}_\nu$ are bounded at zero when $\nu > 0$.

We define a two-parameter family of functions $h_{\nu,n}$, $\nu \in \mathbb{R}$, $n \in \mathbb{Z}_+$, by the following formulas:

$$h_{\nu,0}(t) = \tilde{h}_\nu(t) - h_\nu(t);$$
$$h_{\nu,n}(t) = \frac{\tilde{h}_{\nu+n}(t)}{n!\,2^n} - h_{\nu,n-1}(t), \quad n = 1, 2, \ldots.$$

The functions $h_{\nu,n}$ belong to $\mathcal{M}_{(\lfloor\nu\rfloor+n+1)_+}$ and are bounded at zero when $\nu + n + 1 > 0$. The functions $h_{1-d/2,0}$ with $d = 1, 2$ generate a radial basis of the *tension spline*, and the functions $h_{1-d/2,1}$ with $d = 2, 3$ generate a radial basis of the *regularized spline* in $\mathbb{R}^d$ that were suggested in [30].

Exponential Integral Kernels and Their Generalization

Let us consider a family of functions

$$g_\nu(t) = t^\nu \Gamma(-\nu, t), \quad \nu \in \mathbb{R}.$$

Here $\Gamma(a,t) = \int\limits_t^\infty e^{-x} x^{a-1}\,dx$ is an incomplete gamma function. With $\nu \in \mathbb{Z}$ we obtain the following known exponential integral functions [35, Sections 5.1.45, 5.1.46]:

$$g_n(t) = \begin{cases} \alpha_{-n-1}(t), & n < -1, \\ E_{n+1}(t), & n > -1, \end{cases} \qquad g_{-1}(t) = \alpha_0(t) = E_0(t) = \frac{e^{-t}}{t}.$$

Again, the generalization here consists in extending this representation to arbitrary $\nu \in \mathbb{R}$. In common with h_ν, the functions g_ν belong to $\mathcal{M}_0$ for all $\nu \in \mathbb{R}$ and are bounded at zero with $\nu > 0$.

Generalization of Completely Regularized Splines

Let us consider an auxiliary family of functions

$$\tilde{g}_\nu(t) = \begin{cases} \Gamma(-\nu) t^\nu, & \nu \in \mathbb{R} \setminus \mathbb{Z}_+, \\ (-1)^{\nu+1} \dfrac{t^\nu[\ln t - \psi(\nu+1)]}{\nu!}, & \nu \in \mathbb{Z}_+, \end{cases}$$

which is similar to the family f_ν, and set that

$$g_{\nu,0}(t) = \tilde{g}_\nu(t) - g_\nu(t);$$
$$g_{\nu,n}(t) = g_{\nu+1,n-1}(t) - g_{\nu,n-1}(t), \quad n = 1, 2, \ldots.$$

The functions $g_{\nu,n}$ belong to $\mathcal{M}_{(\lfloor\nu\rfloor+n+1)_+}$ and are bounded at zero at any $\nu \in \mathbb{R}$ and $n \in \mathbb{Z}_+$. The functions $g_{1-d/2,0}$ generate a radial basis of the *completely regularized spline* in $\mathbb{R}^d$ with $d = 2, 3$ suggested in [31].

9. COMPARISON OF RADIAL BASIS FUNCTIONS

When one is choosing a suitable radial basis function and a basis of the trend of an RBF spline, requirements to the smoothness of the solution play an essential role, and besides that, it is important to know the properties of the variational functional $\|f\|_G^2$ minimized by the RBF spline. The table shows the form of variational functional for some known RBF splines with application of the following notation:

$$\|D^k f\|_Y^2 = \sum_{|\alpha|=k} \frac{k!}{\alpha!} \|D^\alpha f\|_Y^2, \quad k \in \mathbb{Z}_+.$$

For example, a thin plate spline [37] that minimizes the norm of the second derivative in $L_2(\mathbb{R}^2)$ is obtained using the function $f_1(r^2)$ ($m = d = 2$). A cubic spline in $\mathbb{R}^1$ is matched by the function $f_{3/2}(r^2)$ ($m = 2,\ d = 1$). The Whittle–Matérn spline that minimizes a norm in Sobolev's space is obtained using the function h_ν. Tension splines and regularized splines are obtained using the function $h_{\nu,n}$ and minimize the weighted sum of the norms of two neighboring derivatives, and the contribution of each of the derivatives can be regulated via the control parameter. There is a special case of the completely regularized spline generated by the function $g_{\nu,0}$. Its native space consists of infinitely differentiable functions. Using a nonzero Hardy parameter c in formula (7.1) also generates a native space of infinitely differentiable functions.

The rate of convergence of splines when the mesh of nodes is condensed is determined by the order of smoothness of the radial basis function at zero. We define a linear operator $\mathcal{T}_\nu$, which relates the function f to the remainder term of its expansion at zero into a Taylor series of degree less than ν:

$$\mathcal{T}_\nu f(t) = f(t) - \sum_{k\in\mathbb{Z}_+,\, k<\nu} \frac{f^{(k)}(0)}{k!}\, t^k.$$

Note that $\mathcal{T}_\nu f = f$ at $\nu \le 0$.

We use the notation $u \approx v$ if $\lim\limits_{t\to 0} \frac{u(t)}{v(t)} = C \ne 0$ and $u \sim v$ if $C = 1$ in this formula. According to [33],

$$\mathcal{T}_\nu h_\nu \sim \tilde{h}_\nu, \quad \mathcal{T}_{\nu+n+1} h_{\nu,n} \sim \frac{\tilde{h}_{\nu+n+1}}{(n+1)!\, 2^{n+1}}, \quad \mathcal{T}_\nu g_\nu \sim \tilde{g}_\nu.$$

Since $f_\nu \approx \tilde{h}_\nu \approx \tilde{g}_\nu$, the functions $h_\nu(r^2)$, $g_\nu(r^2)$, $f_\nu(r^2)$, and $h_{\nu-n-1,n}(r^2)$ generate splines with close convergence properties. Their differences show in the minimum required order m of the polynomial trend:

- $m = 0$ for h_ν and g_ν,
- $m = (\lfloor \nu \rfloor + 1)_+$ for f_ν,

Table

Functional	$\phi(r)$	Restrictions	Spline
$\|D^m f\|_{L_2}^2$	$f_{m-d/2}(r^2)$	$m - d/2 > 0$	Spline in $L_2^m(\mathbb{R}^d)$
$\|D^m f\|_{\tilde{H}^s}^2$	$f_{m+s-d/2}(r^2)$	$m > m+s-d/2 > 0$	Spline in $D^{-m}\tilde{H}^s(\mathbb{R}^d)$ [37]
$\|f\|_{L_2}^2 + \|D^m f\|_{L_2}^2$	$h_{m-d/2}(r^2)$	$m - d/2 > 0$	Spline in $H^m(\mathbb{R}^d)$
$\varphi^2\|D^1 f\|_{L_2}^2 + \|D^2 f\|_{L_2}^2$	$h_{1-d/2,0}((\varphi r)^2)$	$2 - d/2 > 0$	Tension spline [37]
$\|D^2 f\|_{L_2}^2 + \tau^2\|D^3 f\|_{L_2}^2$	$h_{1-d/2,1}((r/\tau)^2)$	$3 - d/2 > 0$	Regularized spline [30]
$\sum\limits_{k=1}^{\infty} \frac{\|D^k f\|_{L_2}^2}{\varphi^{2k}(k-1)!}$	$g_{1-d/2,0}((\varphi r/2)^2)$	—	Completely regularized spline [31]

- $m = (\lfloor \nu \rfloor)_+$ for $h_{\nu-n-1,n}$,

as well as the behavior at infinity: $h_\nu(r^2) \approx r^{\nu-1/2}e^{-r}$, and $g_\nu(r^2) \approx e^{-r^2}/r^2$, while the other functions do not decrease exponentially.

10. PREPARATION OF DATA FOR APPROXIMATION WITH RBF SPLINES

In conclusion, we present recommendations on preparation of data for approximation with RBF splines. When preparing data, it is important to take into account the type of measurement error for the function to approximate: the values can be measured with *absolute error* $z \pm \varepsilon$ or *relative error* $z(1 \pm \varepsilon)$.

Virtually all approximation methods minimize the norm of the approximation absolute error vector, and, thus, if the approximate values were measured with a relative error, then it is necessary to set appropriately the weights in functional (2.6). We can also apply *calibration* of the function to approximate, which converts the relative error into the absolute one, for example $z \mapsto \tilde{z} = \ln z$. After constructing the spline from the calibrated values, we calculated the values of the original function using a reverse transformation, for example $\tilde{z} \mapsto z = \exp \tilde{z}$.

If the values the function to approximate must lie in a certain range, then it is desirable to use a calibration that eliminates the restrictions. For example, a physical process efficiency varying in the range of $[0, 100]$ can be subjected to the calibration $z \mapsto \tilde{z} = \operatorname{tg} \pi(z/100 - 0.5)$.

Independent variables measured with relative error also have to be converted into a form with absolute measurement error. For example, a dependence of the form $S = aP^b$ with the parameters a and b unknown, in which S and P are given with relative error, can be converted to the form $\ln S = \ln a + b \ln P$, in which the variables $\ln P$ and $\ln S$ are measured with absolute error.

Usually, then an affine transformation of the coordinate system is performed with the aim of bringing the independent variables to a common scale and placing the origin within "the cloud" of the points of the mesh. The need to reduce the variables to a common scale is due to the fact that "the influence" of the basis function of RBF spline propagates radially, and the shift of the origin into the cloud of the points of the mesh can significantly affect the accuracy of calculating the basis functions of the spline trend. Of course, the scaling is required only if the independent variables are of different nature.

Usually, *geometric* or *statistical* scaling is applied. When geometric scaling is used, a parallelepiped surrounding the set of the mesh nodes is converted into a cube with the center at the origin. With statistical scaling, independent variables are transformed so that the average value of each coordinate of the nodes of the transformed mesh equals to zero, and the standard deviation is equal to unity. Due this scaling, one can consciously choose parameters of radial basis function. For example, a Hardy parameter $c = 0.01$ will mean 1% of the root-mean-square spread of mesh coordinates in the case of statistical scaling.

If the number of mesh nodes is rather large, then in experiments it would be useful to break them into two parts: a *learning* sample and a *test* one. In this case, experiments on approximation of function are carried out on the learning sample, and the test sample is used in analysis of the quality of approximation (the *cross-check method*). For example, this approach may be useful for scaling a mesh: experimenting with different scaling parameters, we choose them such that the deviation of the interpolating spline constructed on the learning sample is minimal on the test sample.

In problems relating to statistical or medical information, the number of independent variables can be very large and approximation with RBF splines constructed in consideration of all variables has practically zero predictive potential, i.e., it approximates data only in a small neighborhood of the mesh nodes, whereas outside the nodes the accuracy of the approximation tends to zero. In such problems, they try to construct a simplest approximation, for example, using linear regression.

To reduce the complexity of RBF splines, one should first reduce the number of independent variables, for example, via discarding variables that are in weak correlation with the target attribute. Correlation analysis of independent variables, in which only one of several strongly correlated variables is left and the rest are discarded, could also be of use. Finally, the *principal component method* [38] can be applied to reducing the coordinate system of the transformed mesh to the main directions and, secondary directions discarded, decreasing the dimensionality of the space of independent variables. After the reduction in the number of independent variables, it makes sense to use spline regression with selection of datum nodes of spline, for example, using the algorithm described in Section 5.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 15-07-07530.

REFERENCES

1. Schaback, R., Native Hilbert Spaces for Radial Basis Functions. I, in *New Developments in Approximation Theory*, Muller, M.W., Buhmann, M.D., Mache, D.H., and Felten, M., Eds., Basel: Birkhauser, 1999, pp. 255–282.
2. Madych, W.R. and Nelson, S.A., Multivariate Interpolation and Conditionally Positive Definite Functions. II, *Math. Comput.*, 1990, vol. 54, pp. 211–230.
3. Wendland, H., *Scattered Data Approximation*, Cambridge: Cambridge University Press, 2005.
4. Powell, M.J.D., The Theory of Radial Basis Function Approximation, Preprint/DAMTP; NA11, Cambridge, 1990.
5. Schaback, R. and Wendland H., Characterization and Construction of Radial Basis Functions, in *Multivariate Approximation and Applications*, Din, N., Leviatan, D., Levin,D., and Pinkus, A., Eds., Cambridge: Cambridge University Press, 2001, pp. 1–24.
6. Rozhenko, A.I., *Teoriya i algoritmy variatsionnoi splain-approksimatsii* (Theory and Algorithms of Variational Spline Approximation), Novosibirsk: ICMMG SB RAS, 2005.
7. Kovalkov, A.V., On an Algorithm for Constructing Splines with Discrete Inequality Constraints, *SERDIKA, Bolgarsko Matematichesko Spisanie*, 1983, vol. 9, no. 4, pp. 417–424.
8. Rozhenko, A.I. and Fedorov, E.A., On the Algorithm of Smoothing by a Spline with Bilateral Constraints, *Sib. Zh. Vych. Mat.*, 2016, vol. 19, no. 3, pp. 331–342;
9. Atteia, M., Fonctions "Spline" et Noyaux Reproduissants d'Aronszain–Bergman, *RAIRO*, 1970, vol. 4, no. 3, pp. 31–43.
10. Bezhaev, A.Yu., Reproducing Mappings and Vector Spline-Functions, *Sov. J. Num. An. Math. Modell.*, 1990, vol. 5, no. 2, pp. 91–109.
11. Vasilenko, V.A., *Teoriya splain-funktsii* (The Theory of Spline Functions), Novosibirsk: NGU, 1978.
12. Fasshauer, G.E., Green's Functions: Taking Another Look at Kernel Approximation, Radial Basis Functions, and Splines, *Approximation Theory XIII: San Antonio 2010*, Neamtu, M. and Schumaker, L., Eds., Springer, 2012, pp. 37–63.
13. Anselone, P.M. and Laurent, P.J., A General Method for the Construction of Interpolating or Smoothing Spline-Functions, *Num. Math.*, 1968, vol. 12, pp. 66–82.
14. Reinsch, C.H., Smoothing by Spline Functions. II, *Numer. Math.*, 1971, vol. 16, no. 5, pp. 451–454.
15. Gordonova, V.I. and Morozov, V.A., Numerical Algorithms for Selection of Parameters in the Regularization Method, *Zh. Vych. Mat. Mat. Fiz.*, 1973, vol. 13, no. 3, pp. 539–545.
16. Bezhaev, A.Yu. and Vasilenko, V.A., *Variational Spline Theory*, Novosibirsk: NCC, 1993.
17. Rozhenko, A.I., On Optimal Choice of Spline-Smoothing Parameter, *Bull. Novosibirsk Comput. Center. Ser. Num. An.*, Novosibirsk: NCC, 1996, iss. 7, pp. 79–86.
18. Rozhenko, A.I., A New Method for Finding an Optimal Smoothing Parameter of the Abstract Smoothing Spline, *J. Approx. Theory*, 2010, vol. 162, pp. 1117–1127; DOI:10.1016/j.jat.2009.08.002.
19. Mokshin, P.V. and Rozhenko, A.I., On Finding an Optimal Smoothing Parameter, *Sib. Zh. Ind. Mat.*, 2015, vol. 18, no. 2(62), pp. 63–73.
20. Ivanova, E.D., Acceleration of Convergence of Algorithm for Selecting a Smoothing Parameter by the Method of Asymptotic Estimates, Final qualification master's work, Novosibirsk State University, 2011.
21. Rozhenko, A.I., *Abstraktnaya teoriya splainov* (Abstract Theory of Splines), Novosibirsk: NSU, 1999.
22. *Grubbs' Test for Outliers*; https://en.wikipedia.org/wiki/Grubbs'_test_for_outliers.
23. *Netflix Prize*; https://en.wikipedia.org/wiki/Netflix_Prize.
24. Dubrule, O., *Geostatistics for Seismic Data Integration in Earth Models*, Tulsa: Society of Exploration Geophysicists and European Association of Geoscientists and Engineers, 2003.
25. Micchelli, C.A., Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions, *Constr. Approx.*, 1986, vol. 2, pp. 11–22.
26. Duchon, J., Fonctions-Spline à Ènergie Invariante par Rotation, Preprint/RR; 27, Grenoble, 1976.
27. Ignatov, M.I. and Pevnyi, A.B., *Natural'nye splainy mnogikh peremennykh* (Multivariate Natural Splines), Leningrad: Nauka, 1991.
28. Volkov, Yu.S. and Miroshnichenko, V.L., The Construction of a Mathematical Model of Universal Characteristics of a Radial-Axial Turbine, *Sib. Zh. Ind. Mat.*, 1998, vol. 1, no. 1, pp. 77–88.
29. Bogdanov, V.V., Karsten, W.V., Miroshnichenko, V.L., and Volkov, Yu.S., Application of Splines for Determining the Velocity Characteristic of a Medium from a Vertical Seismic Survey, *Central European J. Math.*, 2013, vol. 11, no. 4, pp. 779–786.

30. Mitáš, L. and Mitášová, H., General Variational Approach to the Interpolation Problem, *Comput. Math. Appl.*, 1988, vol. 16, no. 12, pp. 983–992.
31. Mitášová, H. and Mitáš, L., Interpolation by Regularized Splines with Tension: I. Theory and Implementation, *Math. Geol.*, 1993, vol. 25, iss. 6, pp. 641–655.
32. Rozhenko, A.I. and Shaidorov, T.S., On the Construction of Splines by the Reproducing Kernel Method, *Sib. Zh. Vych. Mat.*, 2013, vol. 16, no. 4, pp. 365–376.
33. Rozhenko, A.I., On a New Family of Conditionally Positive Definite Radial Basis Functions, *Tr. IMM UrO RAN*, 2013, vol. 19, no. 2, pp. 256–266.
34. Rozhenko, A.I., New Families of Radial Basis Functions, in *Constructive Theory of Functions*, Ivanov, K., Nikolov, G., and Uluchev, R., Eds., Marin Drinov Academic Publishing House, 2014, pp. 217–233; http://www.math.bas.bg/mathmod/Proceedings_CTF/CTF-2013/files_CTF-2013/17-Rozhenko.pdf.
35. *Spravochnik po spetsial'nym funktsiyam s formulami, grafikami i matematicheskimi tablitsami* (Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables), Abramowitz, M. and Stegun, I.A., Eds., New York: Dover Publications, 1972.
36. Matérn, B., *Spatial Variation*, 2nd ed., Berlin: Springer, 1986.
37. Duchon, J., Spline Minimizing Rotation-Invariant Seminorms in Sobolev Spaces, in *Constructive Theory of Functions of Several Variables*, Berlin: Springer, 1977, pp. 85–100.
38. Principal Component Analysis; https://en.wikipedia.org/wiki/Principal_component_analysis.