Splines and stationary Gaussian processes

Yuri Golubev
CNRS, Université de Provence
Moscow Institute for Information Transmission Problems
PreMoLab
yuri.golubev@gmail.com

Ekaterina Krymova
Moscow Institute for Information Transmission Problems
DATADVANCE
PreMoLab
ekaterina.krymova@datadvance.net

Abstract

Splines are very popular in function interpolation thanks to their robustness and fast computational algorithms. In this talk, it is shown that splines are closely related to the interpolation of stationary Gaussian processes. This fact permits to predict the error of spline interpolation and to compute it very fast. It is shown also that splines are nearly optimal with respect to the minimax interpolation of smooth Gaussian processes and functions from Sobolev's ball.

1. Introduction

In this paper we deal with recovering of an unknown function $f(x)$, $x \in \mathbb{R}$ based on the data $Y_k = f(X_k)$, $k = 1,\ldots,n$.

Among numerous interpolation methods, splines are widely used in practice because they provide a very good interpolation of smooth functions and admit a simple physical interpretation. Apparently, the cubic splines were firstly mentioned in [2].

The main goal in this paper is to show that splines are closely related to the interpolation of stationary Gaussian processes with the spectral density

$$F_\alpha(\omega) = \frac{C}{\omega^{2m} + \alpha^{2m}};$$

where C, α are some positive constants and $m \geq 1$ is a known integer.

Based on this analogy we propose a fast method for computing the accuracy of spline interpolation. It will be shown also splines are nearly optimal in the framework of the minimax interpolation theory of functions and processes from Sobolev's ball.

2. Interpolation of stationary Gaussian process

It is well known (see e. g. [4, 5, 8]) that the best interpolation of the Gaussian process with the spectral density $F_\alpha(\omega)$ is given by

$$\bar{f}(x,Y) = \sum_{k=1}^{n} K_\alpha(x,X_k)Y_k,$$

where $K_\alpha(x,X_k)$ satisfies the Wiener-Hopf equation

$$\mathbf{E}\left[f(x) - \sum_{k=1}^{n} K_\alpha(x,X_k)f(X_k)\right] f(X_s) = 0,\ s = 1,\ldots,n. \quad (1)$$

The interpolation error is computed as follows

$$\begin{aligned}\sigma_\alpha^2(x) &\stackrel{\text{def}}{=} \mathbf{E}\left[f(x) - \sum_{k=1}^{n} K_\alpha(x,X_k)f(X_k)\right]^2 \\ &= \mathbf{E}\left[f(x) - \sum_{k=1}^{n} K_\alpha(x,X_k)f(X_k)\right] f(x).\end{aligned}$$

In practice, the parameters α and C are usually unknown and to make use of (1) one has to estimate them. Usually this is done with the help of the maximum likelihood method. Unfortunately, such an approach complicates greatly the solution to the problem of interpolation. A more robust approach would be to find a solution (1) as $\alpha \to 0$. However, the implementation of this idea is not obvious because

$$\lim_{\alpha\to 0} \frac{C}{\omega^{2m} + \alpha^{2m}} = \frac{C}{\omega^{2m}}$$

and therefore

$$\int_{-\infty}^{\infty} \frac{1}{\omega^{2m}}\, d\omega = \infty.$$

This means that the random process with the spectral density $C\omega^{-2m}$ doesn't exist. Nevertheless, it can be shown that the following limits

$$\lim_{\alpha\to 0} K_\alpha(x,X) = K_0(x,X),$$

$$\lim_{\alpha\to 0} \sigma_\alpha^2(x) = \sigma_0^2(x)$$

do exist. More precisely, it holds

Theorem 1 *Let all X_k, $k = 1,2,\ldots,n$ are distinct and $n \geq m$. Define functions $d_s^{(m)}[x]$ as follows:*

$$\begin{aligned} d_s^{(1)}[x] &= \frac{|X_{s+1} - x|^{2m-1} - |X_s - x|^{2m-1}}{X_{s+1} - X_s}, \\ d_s^{(j+1)}[x] &= \frac{d_{s+1}^{(j)}[x] - d_s^{(j)}[x]}{X_{s+j+1} - X_s}, \end{aligned} \quad (2)$$

where $s = 1,\dots,n-m$. *Then*

$$\lim_{\alpha\to 0} K_\alpha(x,X_k) = K_0(x,X_k),\ k = 1,\dots,n,$$

where $K_0(x,X_k)$ *is a solution to*

$$\sum_{k=1}^{n} K_0(x,X_k) d_s^{(m)}[X_k] = d_s^{(m)}[x], \quad s = 1,\dots,n-m, \tag{3}$$

$$\sum_{k=1}^{n} K_0(x,X_k) X_k^p = x^p, \quad p = 0,\dots,m-1. \tag{4}$$

For the interpolation error we have

$$\lim_{\alpha\to 0} \sigma_\alpha^2(x) = (-1)^{m+1} C \sum_{k=1}^{n} K_0(x,X_k)|x-X_k|^{2m-1}.$$

3. Fast algorithm

From a numerical viewpoint, Equation (3) is not good because its matrix is dense. In this section we show how to transform this equation into a band form and thus to construct a fast algorithm for solving (3).

It follows immediately from (2) and (3) that for $j = m+1,\dots,2m$ the following equations hold

$$\sum_{k=1}^{n} K_0(x,X_k) d_s^{(j)}[X_k] = d_s^{(j)}[x], \quad s = 1,\dots,n-j. \tag{5}$$

It is also easily seen that

$$d_s^{(j)}[x] = P_s^{(2m-j-1)}[x], \quad x \ge X_{s+j},$$
$$d_s^{(j)}[x] = -P_s^{(2m-j-1)}[x], \quad x \le X_s.$$

where $P_s^{(2m-j-1)}[x]$ is the polynomial of order $2m-j-1$ defined by

$$P_s^{(2m-j-1)}[x] = \sum_{q=0}^{2m-j-1} (-1)^{j+q} C_{2m-1}^{j+q} x^{2m-j-q-1} \times \sum_{l_0+\dots+l_j=q} X_s^{l_0} \times \dots \times X_{s+j}^{l_j}.$$

So, $P_s^{(0)}[x] = -1$ and therefore

$$d_s^{(2m)}[x] = 0, \quad x \notin [X_s, X_{s+2m}].$$

Thus with Equations (5) can be rewritten in the following equivalent form

$$\sum_{k=s+1}^{s+2m-1} K_0(x,X_k) d_s^{(2m)}[X_k] = d_s^{(2m)}[x], \tag{6}$$

where $s = 1,\dots,n-2m$.

To obtain the remainder $2m$ equations we combine (5) and (4). Thus we arrive at

$$\sum_{k=1}^{q-1} K_0(x,X_k)\left[d_q^{(q)}[X_k] - P_q^{(2m-1-q)}[X_k]\right] = d_q^{(q)}[x] - P_q^{(2m-1-q)}[x], \quad q = m,\dots,2m-1 \tag{7}$$

and

$$\sum_{k=n-q+1}^{n} K_0(x,X_k)\left[d_{n-q}^{(q)}[X_k] + P_{n-q}^{(2m-1-q)}[X_k]\right] = d_{n-q}^{(q)}[x] + P_{n-q}^{(2m-1-q)}[x], \quad q = 2m-1,\dots,m. \tag{8}$$

It is clear that the matrix of the linear system (7 - 8) has $(2m-1)$ - band form. Notice that in case $m = 2$ the traditional method of solving tridiagonal systems is the Thomas algorithm [1]. When $m > 2$ the Cholesky decomposition [3] may be used to find a solution to (7 - 8).

4. Splines and Gaussian processes

In this section, we consider a slightly more general smoothing problem assuming that we have at our disposal the noisy data

$$Y_j = f_\alpha(X_j) + \varepsilon_j,\ j = 1,\dots,n,$$

where ε_j is a white Gaussian noise with $\mathbf{E}\varepsilon_j^2 = \sigma^2 > 0$ and $f_\alpha(\cdot)$ is a Gaussian process with the spectral density $F_\alpha(\cdot)$. Our goal is to estimate $f(x)$. Let $\bar{f}_\alpha(x,Y)$ be the best smoothing. The following theorem shows that $\lim_{\alpha\to 0} \bar{f}_\alpha(x,Y)$ is a smoothing spline.

Theorem 2 *Let all* $X_j,\ j = 1,2,\dots,n$ *are distinct and* $n \ge m$. *Then*

$$\lim_{\alpha\to 0} \bar{f}^\alpha(x,Y) = \bar{f}(x,Y),$$

where

$$\bar{f}(x,Y) = \arg\min_f \left\{ \frac{1}{2\sigma^2} \sum_{k=1}^{n} \left[Y_j - f(X_j)\right]^2 + \frac{(2\pi)^{2m}}{2C} \int_0^1 \left[f^{(m)}(t)\right]^2 dt \right\}.$$

5. Minimax interpolation of stationary Gaussian processes

Suppose $f(\cdot)$ is a stationary Gaussian process with a known spectral density $F(\omega)$. Assume also that

$$X_k = kh,\ k = 0,\pm 1,\pm 2,\dots.$$

Our goal is to recover the trajectory $f(x)$ on the interval $[0,h]$ based on the data $Y_k = f(X_k),\ k = 0,\pm 1,\pm 2,\dots$. Since the process is Gaussian and stationary it may be easily shown that the best estimate has the following form:

$$\bar{f}(x,Y,K) = h \sum_{k=-\infty}^{\infty} K(x-X_k) Y_k,$$

where $K(\cdot)$ is a symmetric kernel minimizing the interpolation error

$$\sigma^2(K) = \frac{1}{h}\int_0^h \mathbf{E}[f(x) - \bar{f}(x,Y,K)]^2\, dx.$$

Assume that the process $f(x)$ is smooth i.e.

$$\mathbf{E}[f^{(m)}(x)]^2 \leq L. \tag{9}$$

Denote by $\mathscr{F}(m,L)$ the class of all stationary processes for which Condition (9) holds.

Our goal in this section is to compute the minimax interpolation error defined by

$$r_h(m,L) = \inf_{\bar{f}} \sup_{f\in\mathscr{F}(m,L)} \frac{1}{h}\int_0^h \mathbf{E}[f(x)-\bar{f}(x,Y)]^2\,dx.$$

Theorem 3

$$r_h(m,L) = \frac{L}{2}\left(\frac{h}{\pi}\right)^{2m}.$$

Minimax interpolation is given by

$$\bar{f}_\circ(x,Y) = h\sum_{k=-\infty}^{\infty} K_\circ(x-X_k)Y_k,$$

where

$$\hat{K}_\circ(\omega) = \begin{cases} 1, & \omega\in[0,\omega_\circ], \\ 2^{m-1}(1-\omega)^m, & \omega\in[\omega_\circ,1/2], \\ 1-2^{m-1}\omega^m, & \omega[1/2,1-\omega_\circ], \\ 0, & \omega\geq 1-\omega_\circ. \end{cases}$$

and $\omega_\circ = 1-2^{-1+\frac{1}{m}}$.

In order to understand how splines can interpolate random processes from $\mathscr{F}(m,L)$, define the risk of the spline interpolation by

$$r_h^{spline}(m,L) = \sup_{f\in\mathscr{F}^m(L)} \frac{1}{h}\int_0^h \mathbf{E}[f(x)-\bar{f}(x,Y)]^2 dx,$$

where $\bar{f}(x,Y)$ is defined by (2).

Theorem 4

$$r_h^{spline}(m,L) = L\left(\frac{h}{2\pi}\right)^{2m} \max_{\omega}\left\{\frac{1}{\omega^{2m}}\left[[1-\hat{K}_s(\omega)]^2 + \right.\right.$$
$$\left.\left. + \sum_{k\neq 0}\hat{K}_s^2(\omega+k)\right]\right\},$$

where

$$\hat{K}_s(\omega) = \left[1+\sum_{k\neq 0}\left(1+\frac{k}{\omega}\right)^{-2m}\right]^{-1}.$$

Unfortunately, analytic computing of the right-hand side at (4) is rather cumbersome and difficult. Therefore we compute it numerically and trace the minimax spline efficiency $r_h^{spline}(m,L)/r_h(m,L)$. From Figure 1 we see that the risk of spline interpolation is rather close to the minimax risk. Notice that the minimax efficiency of the cubic spline ($m=2$) is approximately 1.35.

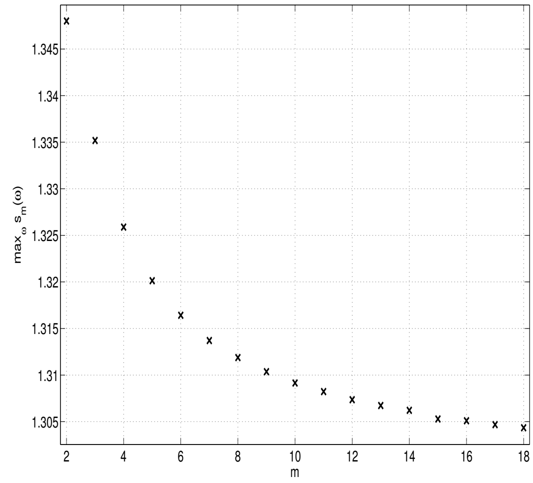

Figure 1. Minimax spline efficiency $r_h^{spline}(m,L)/r_h(m,L)$ for different m.

6. Accuracy evaluation for splines

Usually, in practical applications, along with the interpolation method, one is interested in controlling the accuracy of the interpolation in use. In this section we propose a simple estimate for the spline interpolation accuracy.

Denote by $\bar{f}^m(x,X,Y)$ and $\bar{f}^{m+1}(x,X,Y)$ splines of orders m and $m+1$ respectively. To estimate from above the spline accuracy

$$\sigma^2(x,X,Y) = [f(x)-\bar{f}^m(x,X,Y)]^2,$$

we propose to make use of the following estimate:

$$\bar{\sigma}^2(x,X,Y) = [\bar{f}^{m+1}(x,X,Y)-\bar{f}^m(x,X,Y)]^2.$$

In order to justify this method, we test it on stationary random processes. Assume that $X=\{X_k=kh,\ k=0,\pm1,\pm2,\ldots\}$ and $Y_k=f(X_k)$. The following theorem shows that $\sigma^2(x,X,Y)$ may be controlled by $\bar{\sigma}^2(x,X,Y)$.

Theorem 5 *Suppose $f(x)$ is a stationary process with the spectral density*

$$F(\omega) = \frac{\phi(|\omega|)}{|\omega|^{2m}},$$

where $\phi(\omega)$, $\omega>0$ is a non-negative and non-increasing function. Then there exists a constant Q such, that

$$\int_{X_k}^{X_{k+1}} \mathbf{E}\sigma^2(x,X,Y)\,dx \leq Q\int_{X_k}^{X_{k+1}} \mathbf{E}\bar{\sigma}^2(x,X,Y)\,dx, \tag{10}$$

The constant Q in (10) tends to 1 when the smoothness of the underlying process increases. The next result justifies this assertion.

Theorem 6 *Suppose the spectral density of the random process f has the following form:*

$$F(\omega) = \phi(\omega)\exp(-\mu\omega^2),$$

where $\phi(\cdot)$ fulfills $A\phi_\circ(|\omega|) \le \phi(\omega) \le B\phi_\circ(|\omega|)$ for some non-negative and non-increasing function $\phi_\circ(\cdot)$. Then

$$\int_{x_k}^{x_{k+1}} \mathbf{E}\sigma^2(x,X,Y)\,dx \le Q(\mu)\int_{x_k}^{x_{k+1}} \mathbf{E}\bar{\sigma}^2(x,X,Y)\,dx$$

and

$$\lim_{\mu\to\infty} Q(\mu) = 1.$$

7. Interpolation of smooth functions

Let us now turn to the interpolation of deterministic functions from the Sobolev's ball

$$\mathscr{W}_T^m(L) = \left\{ f: \int_{-\infty}^{\infty} [f^{(m)}(x)]^2\,dx \le LT, \right.$$
$$\left. \operatorname{supp}\{f\} \in [0,T] \right\}.$$

In what follows it is assumed that the design points X_k are located on the equidistant grid with the step $h > 0$

$$X_k^\zeta = kh + \zeta,\ k = 0 \pm 1, \ldots,$$

where ζ is a random variable uniformly distributed on $[0,h]$.

Our goal is to recover $f(x),\ x \in [0,T]$ based on the data $Y_k^\zeta = f(X_k^\zeta)$. Define the minimax risk

$$\rho_h(m,L) = \lim_{T\to\infty} \inf_{\tilde{f}} \sup_{f\in\mathscr{W}_T^m(L)} \mathbf{E}_\zeta \frac{1}{T}\int_0^T [f(x) - \tilde{f}(x,Y^\zeta)]^2\,dx,$$

where $\mathbf{E}_\zeta$ is averaging with respect to ζ.

The next result shows that the minimax interpolations of functions and Gaussian random processes are very close provided that the design points belong to the randomly positioned equidistant grid.

Theorem 7 *The following equality holds*

$$\rho_h(m,L) = \frac{L}{2}\left(\frac{h}{\pi}\right)^{2m}.$$

The proof of this theorem is based on the method proposed by M. Pinsker in [7].

8. Numerical experiment

In this section we compare numerically the cubic spline interpolation with the kriging (package DACE [6]). The following test functions were used for this comparison:

1. Discontinuous function
$$f(x) = \Big[\mathbf{1}\{0.25 \le x \le 0.75\} + 0.25\Big]\cos(3\pi x).$$
2. Smooth function
$$f(x) = \exp\big(-270|3x-1.8|^3\big) + \exp\big(-350|3x-1.3|^3\big).$$
3. Oscillating function with increasing frequency
$$f(x) = \cos\big[2\pi(0.2x + 7x^2 + 1)\big].$$

To compare the quality of the interpolation and the accuracy evaluation, we begin with computing interpolations for the cubic spline and the kriging using the data $Y_k = f(X_k),\ X_k = k/n,\quad k = 0,\ldots,n$ for a given function f. Next we calculate the actual interpolation error and the predicted error on a very fine grid $x_j = j/N,\quad 1,\ldots,N,\ N \gg n$. Denote these errors by $E_{actual}^{spline}(x,f), E_{pred}^{spline}(x,f)$ and $E_{actual}^{kriging}(x,f), E_{pred}^{kriging}(x,f)$.

We estimate the quality of the interpolation method by means of the mixture of the actual and predicted errors, i.e.

$$E_p^{spline}(x,f) = pE_{pred}^{spline}(x,f) + (1-p)E_{actual}^{spline}(x,f),$$
$$E_p^{kriging}(x,f) = pE_{pred}^{kriging}(x,f) + (1-p)E_{actual}^{kriging}(x,f)$$

where $p \in [0,1]$. Finally, we compare the interpolation

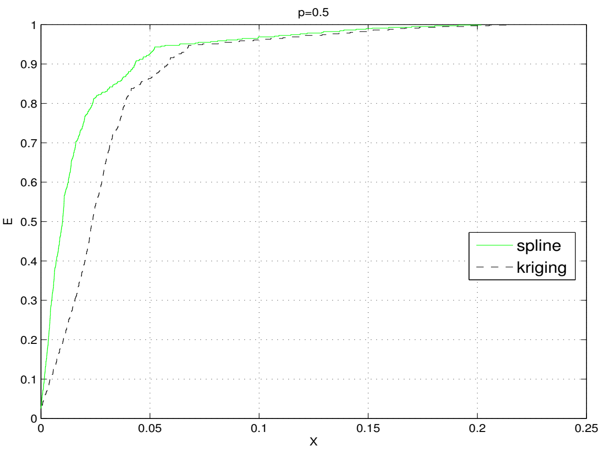

Figure 2. The empirical distribution function of the combined errors in case of the discontinuous function

methods by tracing the empirical distribution functions of the mixture errors $E_p^{spline}(x,f)$ and $E_p^{kriging}(x,f)$ i.e.

$$F_p^{kriging}(z,f) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\Big\{E_p^{kriging}(x,f) \le z\Big\},$$
$$F_p^{spline}(z,f) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\Big\{E_p^{spline}(x,f) \le z\Big\}.$$

On Figures 2-4 the empirical distribution functions are presented for $p = 0.5$. Notice that the larger empirical distribution function means the smaller combination of actual and predicted interpolation errors.

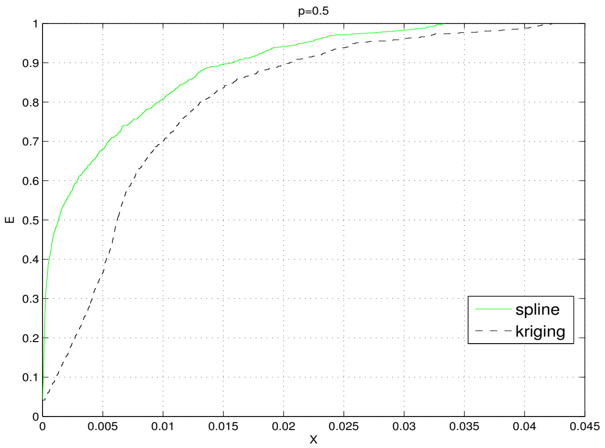

Figure 3. The empirical distribution function of the combined errors in the case of the smooth function

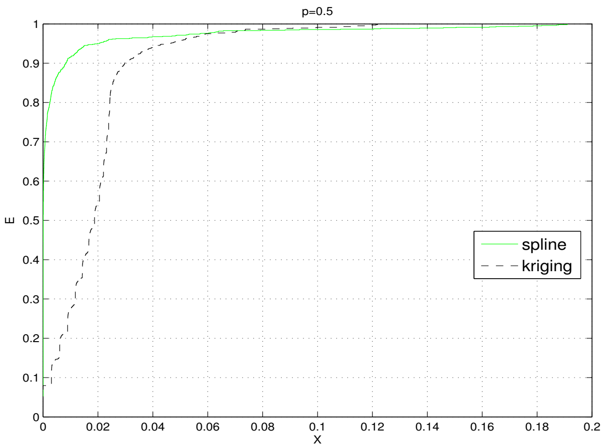

Figure 4. The empirical distribution function of the combined errors in case of the oscillating function.

9. Conclusion

Using the fact that splines may be viewed as limits of interpolations related to stationary Gaussian processes, we propose a very simple and fast method for controlling the accuracy of spline interpolation. It is proved also that splines are nearly optimal with respect to the minimax interpolation theory for smooth processes and functions. Our numerical experiment shows that the proposed accuracy evaluation for splines may be more effective compared with the kriging.

The authors were partially supported by Laboratory of Structural Methods of Predictive Modeling and Optimization, MIPT, RF government grant, ag. 11.G34.31.0073.

References

[1] S.D. Conte and C. deBoor, *Elementary Numerical Analysis*, McGraw-Hill, New York, 1972.

[2] H.-L. Duhamel du Monceau, *Eléments de l'Architecture Navale ou Traité de la Construction des Vaissaux*, Paris, 1752.

[3] G. H. Golub and C. F. Van Loan, *Matrix computations*, Johns Hopkins University Press, ISBN 0-8018-5414-8.

[4] A. N. Kolmogorov, *Stationary sequences in Hilbert space*, Byul. Moskov. Gos. Univ. Ser. Mat. (1941), vol. 2, n. 6, 3–40.

[5] A.N. Kolmogorov, *Interpolation and Extrapolation of stationary random sequences*, Izv. Akad. Nauk SSSR Ser. Mat. (1941), vol. 5, 3–14

[6] S.N. Lophaven, H.B. Nielsen and J. Søndergaard, *DACE - A Matlab Kriging Toolbox, Version 2.0. Report IMM-REP-2002-12*, Informatics and Mathematical Modelling, DTU, 2002.

[7] M. S. Pinsker, *Optimal Filtering of Square-Integrable Signals in Gaussian Noise*, Probl. Peredachi Inf. (1980), 52–68.

[8] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Wiley, New York, 1949.