PADOVA UNIVERSITY PRESS

Dolomites Research Notes on Approximation

Volume 10 · 2017 · Pages 43–50

An Explicit Example of Leave-One-Out Cross-Validation Parameter Estimation for a Univariate Radial Basis Function

L. Bos[a] · F. Polato[b]

Communicated by S. De Marchi

Abstract

We give an explicit example for the selection of the shape parameter for a certain univariate radial basis function (RBF) interpolation problem.

1 Introduction

Radial Basis Function interpolation (RBF) is an important method of (multivariate) interpolation of typically scattered data, which has been much used in applications. The *basic* form of RBF is as follows. Given a *basis* function $g:\mathbb{R}^+\to\mathbb{R}$, the associated RBF interpolant of a data set $\{(x_j,y_j)\}\subset\mathbb{R}^{d+1}$ with n "sites" $x_j\in\mathbb{R}^d$ and function values $y_j\in\mathbb{R}$, is the function of the form $s(x)=\sum_{j=1}^n a_i g(|x-x_j|)$ such that $s(x_i)=y_i$, $1\le i\le n$ (if it exists). Typically the basis function g has a so-called *shape* parameter, the value of which has an important effect on both the quality of the resulting interpolant as well as the numerical conditioning of associated interpolation linear system to be solved. For example, for the Gaussian basis function

$$g_\lambda(x):=\exp(-\lambda\|x\|_2^2),\ \lambda>0$$

small $\lambda\approx 0$ gives a basis function nearly constant in a neighbourhood of the origin, while large $\lambda\approx\infty$ gives a basis function, for all intents, a delta function supported at the origin.

A discussion of practical methods for choosing the shape parameter may be found, for example, in [5, Chapt. 17], where it may be verified that the problem of selecting the shape parameter is indeed important and, in general, rather difficult. One may also consult the monographs [4, 7] for more on the theory of RBF.

Given this typical difficulty of analyzing multivariate interpolation procedures, it is often useful to look more carefully at the univariate case for some suggestion as to how the general case might behave. The goal of this paper is to give an explicit univariate example of one of the most commonly used procedures for selecting an "optimal" shape parameter, the so-called Leave-One-Out Cross-Validation procedure, in the hope that it sheds some light on what happens more generally.

2 Leave-One-Out Cross-Validation (LOOCV)

Consider RBF interpolation with a basis function $g_\lambda:\mathbb{R}^+\to\mathbb{R}$ dependent on some shape parameter λ. For a collection of sites $X=\{x_1,\ldots,x_n\}\subset\mathbb{R}^d$ we let

$$X_j:=X\backslash\{x_j\},\ j=1,2,\ldots,n$$

i.e., X_j is the set of sites with x_j left out. Then, let

$$s_j(x)=\sum_{k\neq j}a_k^{(j)}g_\lambda(|x-x_k|)$$

such that

$$s_j(x_k)=y_k,\ k\neq j.$$

In other words, s_j is the RBF interpolant for the sites X_j. We may think of the value $s_j(x_j)$ as the predicted value of the data at X_j for the left out site x_j, and $e_j:=y_j-s_j(x_j)$, $1\le j\le n$, measures its discrepancy with the value in the full dataset.

LOOCV selects the parameter λ to minimize the 2-norm of the vector of discrepancies e_j, i.e., to minimize

$$E(\lambda):=\sum_{j=1}^n|e_j|^2.$$

[a]Dept. of Computer Science, University of Verona, Italy
[b]Dept. of Computer Science, University of Verona, Italy

Our example will be the LOOCV procedure applied to the functions

$$g_\lambda(x) := ae^{\lambda x} + be^{-\lambda x},\ \lambda, a, b \in \mathbb{R} \tag{1}$$

and sites

$$0 = x_1 < x_2 < x_3 < \ldots < x_n = 1. \tag{2}$$

RBF interpolation by such g_λ was considered in [1] where it is shown (Theorem 1) that the determinant of the associated interpolation matrix is

$$\det\left([g_\lambda(|x_i - x_j|)]_{1\le i,j\le n}\right) = (b-a)^{n-2} e^{-2\lambda\sum_{j=1}^n x_j}\left(\prod_{j=1}^{n-1}(e^{2\lambda x_{j+1}} - e^{2\lambda x_j})\right)\left(b^2 e^{2\lambda x_1} - a^2 e^{2\lambda x_n}\right).$$

Naturally, we must restrict the values of $a, b \in \mathbb{R}$ so that this determinant is non-zero and hence the interpolation problem has a unique solution.

Given this restriction, it is then shown in [1, Theorem 5] that the cardinal functions, u_k (i.e., those linear combinations of the $g_\lambda(|\cdot - x_i|)$ with the property that $u_k(x_i) = \delta_{ik}$) are given by

$$u_k(x) = e^{\lambda(x_k - x)}\begin{cases} \frac{e^{2\lambda x} - e^{2\lambda x_{k-1}}}{e^{2\lambda x_k} - e^{2\lambda x_{k-1}}} & \text{if } x \in [x_{k-1}, x_k] \\ \frac{e^{2\lambda x} - e^{2\lambda x_{k+1}}}{e^{2\lambda x_k} - e^{2\lambda x_{k+1}}} & \text{if } x \in [x_k, x_{k+1}] \\ 0 & \text{otherwise}\end{cases},\ 2 \le k \le n-1,$$

$$u_1(x) = e^{\lambda(x_1 - x)}\begin{cases} \frac{e^{2\lambda x} - e^{2\lambda x_2}}{e^{2\lambda x_1} - e^{2\lambda x_2}} & \text{if } x \in [x_1, x_2] \\ 0 & \text{otherwise}\end{cases},$$

$$u_n(x) = e^{\lambda(x_n - x)}\begin{cases} \frac{e^{2\lambda x} - e^{2\lambda x_{n-1}}}{e^{2\lambda x_n} - e^{2\lambda x_{n-1}}} & \text{if } x \in [x_{n-1}, x_n] \\ 0 & \text{otherwise}\end{cases}$$

independently of the values of a and b!

This allows us, for convenience's sake, to take $b = -a$ with $a = 1/(2\lambda)$ for which our basis function (1) becomes

$$g_\lambda(x) = \frac{\sinh(\lambda x)}{\lambda}.$$

Note that

$$\lim_{\lambda\to 0} g_\lambda(x) = x$$

and the interpolation becomes one by piecewise linear functions. We will therefore let $g_0(x) := x$.

Remark 1. It is worth noting at this point that as λ increases the cardinal functions become more and more like delta functions. Indeed, it is easy to verify that

$$\lim_{\lambda\to\infty} u_k(x) = \begin{cases} 1 & \text{for } x = x_k \\ 0 & \text{for } x \ne x_k. \end{cases}$$

We will actually set $x_1 = 0$ and $x_n = 1$ and select the parameter λ according to the LOOCV principle of minimizing

$$E(\lambda) := \sum_{j=2}^{n-1} |e_j|^2$$

where

$$e_j := s_j(x_j) - y_j,\ 2 \le j \le n-1$$

and s_j is the interpolant of the sites X_j, explicitly

$$0 = x_0 < x_1 < \ldots < x_{j-1} < x_{j+1} < \ldots < x_n = 1. \tag{3}$$

To calculate the values $s_j(x_j)$ we will use the formulas for the cardinal functions u_k given above. Indeed, let $u_k^{(j)}$ be the kth cardinal function for the sites (3). Then

$$s_j(x) = \sum_{k=0, k\ne j}^{n} y_k u_k^{(j)}(x)$$

and, in particular, due to the compact support of the cardinal functions,

$$s_j(x_j) = y_{j-1}u_{j-1}^{(j)}(x_j) + y_{j+1}u_{j+1}^{(j)}(x_j).$$

It is easily verified that then

$$s_j(x_j) = y_{j-1}e^{-\lambda h_{j-1}}\frac{e^{2\lambda h_j} - 1}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} + y_{j+1}e^{\lambda h_j}\frac{1 - e^{-2\lambda h_{j-1}}}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}}$$

where, as usual, we have set $h_j := x_{j+1} - x_j$, $1 \le j \le n-1$. It follows that

$$e_j := y_j - \left\{ y_{j-1} e^{-\lambda h_{j-1}} \frac{e^{2\lambda h_j} - 1}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} + y_{j+1} e^{\lambda h_j} \frac{1 - e^{-2\lambda h_{j-1}}}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} \right\} \tag{4}$$

and

$$E(\lambda) = \sum_{j=2}^{n-1} \left(y_{j-1} e^{-\lambda h_{j-1}} \frac{e^{2\lambda h_j} - 1}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} + y_{j+1} e^{\lambda h_j} \frac{1 - e^{-2\lambda h_{j-1}}}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} - y_j \right)^2 . \tag{5}$$

We note that, since the interchange of a and b in the definition (1) of g_λ is equivalent to replacing λ by $-\lambda$, the formulas for u_k are invariant under this replacement, λ by $-\lambda$. Consequently, $E(-\lambda) = E(\lambda)$ and $E(\lambda)$ is an *even* function.

Remark 2. In case the data comes from $y(x) = \sinh(\alpha x)$ for some $\alpha \in \mathbb{R}$, we note that then

$$y(x) = \alpha g_\alpha(x) = \alpha g_\alpha(|x - x_1|), \; x \in [0,1],$$

i.e., y is in the span of the translates $g(|\cdot - x_j|)$ and so its interpolant is itself. Consequently $E(\alpha) = 0$ and $\lambda = \alpha$ is the optimal parameter. Further, as the $u_k(x)$ do *not* depend on the constants a, b in (1), we also have, for example, that the optimal $\lambda = \alpha$ for $y(x) = \exp(\alpha x)$.

Remark 3. The optimal λ need not be unique. For example, if we take $n = 3$ with $y_1 = y_3 = 0$ then $E(\lambda) = (0 - y_2)^2$ is constant in λ.

Remark 4. An optimal λ may not exist. For example, if we take $n = 3$ with $y_1 = y_3 = +1$ and $y_2 = -1$ then

$$E(\lambda) = \left(e^{-\lambda h_1} \frac{e^{2\lambda h_2} - 1}{e^{2\lambda h_2} - e^{-2\lambda h_1}} + e^{\lambda h_2} \frac{1 - e^{-2\lambda h_1}}{e^{2\lambda h_2} - e^{-2\lambda h_1}} + 1 \right)^2 .$$

As the terms are all positive $E(\lambda) > 1$ while, as is easily verified, $\lim_{\lambda \to \infty} E(\lambda) = 1$ (cf. Remark 1).

2.1 The case of a positive concave function

Theorem 2.1. *Suppose that $y(x) \ge 0$ is concave ($y''(x) \le 0$) for $x \in [0,1]$. Then $\lambda = 0$ is an optimal LOOCV parameter, for any set of sites (2).*

Proof. The discrepancy at x_j is given by e_j, (4). For $\lambda = 0$ this becomes

$$e_j^{(0)} := y_j - \left\{ y_{j-1} \frac{h_j}{h_{j-1} + h_j} + y_{j+1} \frac{h_{j-1}}{h_{j-1} + h_j} \right\}. \tag{6}$$

We claim that for $2 \le j \le (n-1)$,

$$e_j \ge e_j^{(0)} \; (\ge 0 \text{ since } y(x) \text{ is concave}).$$

To see this, first note that

$$\begin{aligned}
& e_j \ge e_j^{(0)} \\
\iff & y_j - \left\{ y_{j-1} e^{-\lambda h_{j-1}} \frac{e^{2\lambda h_j} - 1}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} + y_{j+1} e^{\lambda h_j} \frac{1 - e^{-2\lambda h_{j-1}}}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} \right\} \\
& \ge y_j - \left\{ y_{j-1} \frac{h_j}{h_{j-1} + h_j} + y_{j+1} \frac{h_{j-1}}{h_{j-1} + h_j} \right\} \\
\iff & y_{j-1} \frac{h_j}{h_{j-1} + h_j} + y_{j+1} \frac{h_{j-1}}{h_{j-1} + h_j} \\
& \ge y_{j-1} e^{-\lambda h_{j-1}} \frac{e^{2\lambda h_j} - 1}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} + y_{j+1} e^{\lambda h_j} \frac{1 - e^{-2\lambda h_{j-1}}}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}}.
\end{aligned}$$

Since, by assumption, $y_{j\pm1} \ge 0$, for this it suffices to show that

$$\frac{h_j}{h_{j-1} + h_j} \ge e^{-\lambda h_{j-1}} \frac{e^{2\lambda h_j} - 1}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} \tag{7}$$

and

$$\frac{h_{j-1}}{h_{j-1} + h_j} \ge e^{\lambda h_j} \frac{1 - e^{-2\lambda h_{j-1}}}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}}. \tag{8}$$

To see (7). set $x := \lambda h_{j-1}$ and $y := \lambda h_j$. Then

$$\frac{h_j}{h_{j-1} + h_j} = \frac{y}{x + y}$$

while

$$\begin{aligned} e^{-\lambda h_{j-1}} \frac{e^{2\lambda h_j} - 1}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} &= \frac{e^{x+2y} - e^x}{e^{2(x+y)} - 1} \\ &\le \frac{y}{x+y} \quad \text{(by Lemma 2.2 below)} \\ &= \frac{h_j}{h_{j-1} + h_j}. \end{aligned}$$

Similarly, for (8), set $x := \lambda h_j$ and $y := \lambda h_{j-1}$ so that

$$\begin{aligned} e^{\lambda h_j} \frac{1 - e^{-2\lambda h_{j-1}}}{e^{2\lambda h_j} - e^{-2\lambda h_{j-1}}} &= \frac{e^{x+2y} - e^x}{e^{2(x+y)} - 1} \\ &\le \frac{y}{x+y} \quad \text{(by Lemma 2.2 below)} \\ &= \frac{h_{j-1}}{h_{j-1} + h_j}. \end{aligned}$$

□

Lemma 2.2. *For all $x, y > 0$,*

$$\frac{e^{2y+x} - e^x}{e^{2(x+y)} - 1} \le \frac{y}{x+y}.$$

Proof. This holds iff

$$\begin{aligned} \frac{e^{2y+x} - e^x}{y} &\le \frac{e^{2(x+y)} - 1}{x+y} \\ \iff \quad e^x \frac{e^{2y} - 1}{y} &\le \frac{e^{2(x+y)} - 1}{x+y} \\ \iff \quad \frac{e^{2y} - 1}{y} &\le e^{-x} \frac{e^{2(x+y)} - 1}{x+y} \\ \iff \quad h(y) &\le e^{-x} h(x+y) \end{aligned}$$

for $h(t) := (e^{2t} - 1)/t$.

Hence consider, for a fixed $y > 0$,

$$f(x) := e^{-x} h(x+y).$$

We need to show that $f(x) \ge h(y) = f(0)$, $x \ge 0$, i.e., that the minimum of f on $[0, \infty)$ is $f(0)$. To see this we calculate

$$\begin{aligned} f'(x) &= -e^{-x} h(x+y) + e^{-x} h'(x+y) \\ &= e^{-x}\{h'(x+y) - h(x+y)\}. \end{aligned}$$

But by Lemma 2.3, $h'(x+y) \ge h(x+y)$ and so $f'(x) \ge 0$ and f is increasing on $[0, \infty)$. □

Lemma 2.3. *Let*

$$h(t) := \frac{e^{2t} - 1}{t}$$

(with $h(0) := \lim_{t \to 0} h(t) = 2$). Then, for $t \ge 0$, $h'(t) \ge h(t)$.

Proof. We calculate

$$h'(t) = \frac{2te^{2t} - e^{2t} + 1}{t^2}.$$

Hence $h'(t) \ge h(t)$

$$\begin{aligned} \iff \quad & \frac{(2t-1)e^{2t} + 1}{t^2} \ge \frac{e^{2t} - 1}{t} \\ \iff \quad & (2t-1)e^{2t} + 1 \ge t(e^{2t} - 1) \\ \iff \quad & (t-1)e^{2t} + 1 \ge -t \\ \iff \quad & (t-1)e^{2t} + 1 + t \ge 0. \end{aligned}$$

Now, if $t \ge 1$, $(t-1) \ge 0$ and this latter inequality is clearly true. On the other hand, if $0 \le t < 1$, then $(t-1) < 0$, and

$$\begin{aligned} & (t-1)e^{2t} + 1 + t \ge 0 \\ \iff \quad & 1 + t \ge (1-t)e^{2t} \\ \iff \quad & \frac{1+t}{1-t} \ge e^{2t}, \end{aligned}$$

which is true by Lemma 2.4. □

Lemma 2.4. *For* $0 \le t < 1$,

$$e^{2t} \le \frac{1+t}{1-t}.$$

Proof. The Taylor series for e^{2t} is

$$e^{2t} = 1 + \sum_{k=1}^{\infty} \frac{2^k}{k!} t^k.$$

while the Taylor series for $(1+t)/(1-t)$ is

$$\begin{aligned}\frac{1+t}{1-t} &= \frac{2}{1-t} - 1\\ &= 2\sum_{k=0}^{\infty} t^k - 1\\ &= 1 + \sum_{k=1}^{\infty} 2t^k.\end{aligned}$$

Now it is easy to confirm by induction that $2^k/k! \le 2$, $k = 1, 2, \ldots$. Hence, comparing Taylor series, we are done. □

2.2 The case of equally spaced sites

Here we consider the sites $x_j := (j-1)h$, $1 \le j \le n$ for $h := 1/(n-1)$. In this the e_j simplify to

$$e_j = y_j - \frac{1}{e^{\lambda h} + e^{-\lambda h}} \left\{y_{j-1} + y_{j+1}\right\} = y_j - \frac{1}{2}\mathrm{sech}(\lambda h) \left\{y_{j-1} + y_{j+1}\right\}$$

and

$$E(\lambda) = \sum_{j=2}^{n-1} \left(y_j - \frac{1}{2}\mathrm{sech}(\lambda h) \left\{y_{j-1} + y_{j+1}\right\}\right)^2.$$

Since, as noted previously, $E(-\lambda) = E(\lambda)$ we minimize over the interval $[0, \infty)$. We easily calculate

$$\begin{aligned}E'(\lambda) &= h\,\mathrm{sech}(\lambda h)\tanh(\lambda h)\\ &\quad \times \sum_{j=2}^{n-1} \left(y_j - \frac{1}{2}\mathrm{sech}(\lambda h) \left\{y_{j-1} + y_{j+1}\right\}\right) \left\{y_{j-1} + y_{j+1}\right\}\\ &= h\,\mathrm{sech}(\lambda h)\tanh(\lambda h)\\ &\quad \times \left\{ \sum_{j=2}^{n-1} y_j(y_{j-1} + y_{j+1}) - \frac{1}{2}\mathrm{sech}(\lambda h) \sum_{j=2}^{n-1} (y_{j-1} + y_{j+1})^2 \right\}\\ &= h\,\mathrm{sech}(\lambda h)\tanh(\lambda h) \left(A - \frac{1}{2}\mathrm{sech}(\lambda h) B\right)\end{aligned}$$

where we have set

$$A := \sum_{j=2}^{n-1} y_j(y_{j-1} + y_{j+1}) \text{ and } B := \sum_{j=2}^{n-1} (y_{j-1} + y_{j+1})^2.$$

First note that $h\,\mathrm{sech}(\lambda h)\tanh(\lambda h) = 0$ iff $\lambda = 0$ which is already an endpoint of our interval. The case of $B = 0$ is a bit special. For then, for $\lambda > 0$, $\mathrm{sgn}(E'(\lambda)) = \mathrm{sgn}(A)$. Hence, then, $\lambda = 0$ is the minimum if $A > 0$, there is no minimum if $A < 0$ and $E(\lambda)$ is constant if $A = 0$.

Suppose then that $B \ne 0$. Then, as $B > 0$ and $\mathrm{sech}(t) \le 1$, if $A \ge B/2$ then $E'(\lambda) > 0$ for $\lambda > 0$ and $\lambda = 0$ is the unique optimal parameter.

In case $A \le 0$ then $E'(\lambda) < 0$ for $\lambda > 0$ and again there is no minimum on $[0, \infty)$.

Otherwise, in case $0 < A < B/2$ then there is a critical point given by

$$\mathrm{sech}(\lambda h) = 2A/B,\ \lambda = \frac{1}{h}\mathrm{sech}^{-1}(2A/B)$$

which must necessarily be the optimum λ.

If we let $n \to \infty$ we may observe that for $y(x) \in C^2[0,1]$,

$$\begin{aligned}
\frac{2A}{B} &= \frac{2\sum_{j=2}^{n-1} y_j(y_{j-1}+y_{j+1})}{\sum_{j=2}^{n-1}(y_{j-1}+y_{j+1})^2} \\
&= \frac{2h\sum_{j=2}^{n-1} y_j(y_{j-1}+y_{j+1})}{h\sum_{j=2}^{n-1}(y_{j-1}+y_{j+1})^2} \\
&= \frac{4\int_0^1 (y(x))^2dx + 2h^2\int_0^1 y(x)y''(x)dx + O(h^3)}{4\int_0^1 (y(x))^2 + 4h^2\int_0^1 y(x)y''(x)dx + O(h^3)} \\
&= \frac{1+\frac{1}{2}h^2\int_0^1 y(x)y''(x)dx/\int_0^1 (y(x))^2dx + O(h^3)}{1+h^2\int_0^1 y(x)y''(x)dx/\int_0^1 (y(x))^2dx + O(h^3)} \\
&= 1 - \frac{h^2}{2}\frac{\int_0^1 y(x)y''(x)dx}{\int_0^1 (y(x))^2dx} + O(h^3).
\end{aligned}$$

In the case that

$$\int_0^1 y(x)y''(x)dx \geq 0$$

this expression is at most 1 (for h sufficiently large). Assuming this to be the case and using the fact that

$$\operatorname{sech}^{-1}(t) = \log\left(\frac{1+\sqrt{1-t^2}}{t}\right),$$

we obtain that, for large n the optimal parameter is then

$$\lambda = \sqrt{\frac{\int_0^1 y(x)y''(x)dx}{\int_0^1 (y(x))^2dx}} + O(h).$$

3 Comparison with a Maximum Likelihood Estimate

In the context of Kriging, in which RBF interpolation is embedded in a statistical context, it has been suggested to use a Maximum Likelihood Estimate (MLE) for the optimal parameter. Here, the interpolation matrix $R \in \mathbb{R}^{n\times n}$ given by $R = [g_\lambda(|x_i - x_j|)]_{1\leq i,j\leq n}$, is interpreted as a covariance matrix for a certain family of n random variables. As such R is assumed to be positive definite, and then the MLE parameter turns out be the λ (see e.g. [6]) for which

$$m(\lambda) := |\det(R)|^{1/n}|y^tR^{-1}y| \tag{9}$$

is a *minimum*.

In the case of our example, the interpolation matrix is *not* positive definite (it is however conditionally definite on a $(n-1)$-dimensional subspace) and hence the statistical interpretation of Kriging does not directly apply. However, one may nevertheless attempt to minimize the expression (9) and compare with the LOOCV parameter. Indeed doing so reveals an interesting relation between the two approaches. We concentrate on the case of equally spaced sites, $x_j = (j-1)h$, $1 \leq j \leq n$, $h := 1/(n-1)$.

First note that from Propositions 3.1 and 3.2 of [1] we have then that

$$R^{-1} = \frac{\lambda}{2\sinh(h\lambda)} \times$$
$$\begin{bmatrix}
-\frac{\sinh((1-h)\lambda)}{\sinh(\lambda)} & 1 & 0 & \cdots & 0 & \frac{\sinh(h\lambda)}{\sinh(\lambda)} \\
1 & -2\cosh(h\lambda) & 1 & 0 & \cdots & 0 \\
0 & 1 & -2\cosh(h\lambda) & 1 & \cdots & 0 \\
\cdot & \cdot & & & & \cdot \\
\cdot & \cdot & & & & \cdot \\
0 & \cdots & 0 & 1 & -2\cosh(h\lambda) & 1 \\
\frac{\sinh(h\lambda)}{\sinh(\lambda)} & 0 & \cdots & 0 & 1 & -\frac{\sinh((1-h)\lambda)}{\sinh(\lambda)}
\end{bmatrix}.$$

Let $M := \dfrac{2\sinh(h\lambda)}{\lambda}R^{-1}$ be the above matrix. We caluclate

$$\begin{aligned}
y^tMy &= y_1\left\{-\frac{\sinh((1-h)\lambda)}{\sinh(\lambda)}y_1 + y_2 + \frac{\sinh(h\lambda)}{\sinh(\lambda)}y_n\right\}\\
&\quad+\sum_{i=2}^{n-1} y_i\left\{y_{i-1} - 2\cosh(h\lambda)y_i + y_{i+1}\right\}\\
&\quad+y_n\left\{\frac{\sinh(h\lambda)}{\sinh(\lambda)}y_1 + y_{n-1} - \frac{\sinh((1-h)\lambda)}{\sinh(\lambda)}y_n\right\}\\
&= hy_1\left\{\frac{y_2-y_1}{h} + \frac{1}{h}\left(1-\frac{\sinh((1-h)\lambda)}{\sinh(\lambda)}\right)y_1 + \frac{\sinh(h\lambda)}{h\sinh(\lambda)}y_n\right\}\\
&\quad+h^2\sum_{i=2}^{n-1}\left\{y_i\frac{y_{i-1}-2y_i+y_{i+1}}{h^2} + 2y_i^2\left(\frac{1-\cosh(h\lambda)}{h^2}\right)\right\}\\
&\quad+hy_n\left\{\frac{\sinh(h\lambda)}{h\sinh(\lambda)}y_1 + \frac{y_{n-1}-y_n}{h} + \frac{1}{h}\left(1-\frac{\sinh((1-h)\lambda)}{\sinh(\lambda)}\right)y_n\right\}.
\end{aligned}$$

Then taking the limit as $h \to 0^+$, we see that

$$\begin{aligned}
\lim_{h\to 0^+}\frac{1}{h}y^tMy &= y(0)\left(y'(0) + \lambda\coth(\lambda)y(0) + \frac{\lambda}{\sinh(\lambda)}y(1)\right)\\
&\quad+\int_0^1 y(x)y''(x)dx - \lambda^2\int_0^1 y^2(x)dx\\
&\quad+y(1)\left(-y'(1) + \lambda\coth(\lambda)y(1) + \frac{\lambda}{\sinh(\lambda)}y(0)\right).
\end{aligned}$$

Now, notice that $m(\lambda)$ being a positive value will be minimized if for some λ, $y^tR^{-1}y = 0$, or, equivalently, $y^tMy = 0$. If we were to ignore the boundary terms involving the values of $y(x)$ and $y'(x)$ at $x = 0, 1$ this would happen (approximately) when

$$\int_0^1 y(x)y''(x)dx - \lambda^2\int_0^1 y^2(x)dx = 0,$$

i.e., for

$$\lambda = \sqrt{\frac{\int_0^1 y(x)y''(x)dx}{\int_0^1 (y(x))^2 dx}},$$

i.e., for essentially the *same* value as the LOOCV parameter in this circumstance!

However, the boundary terms do alter this optimal value of the parameter. In Figure 1 below we give the plots of the resulting interpolants for the LOOCV parameter ($\lambda = 1.000658$) and the MLE estimate computed numerically ($\lambda = 2.563091$) and $n = 13$. The MLE interpolant is noticably worse. However, we emphasize that as R is not positive definite the MLE approach is technically not applicable. It is nontheless interesting, that apart from the boundary effects the two approaches provide the same parameter estimate, in this circumstance.

4 Conclusions

We have given a univariate example where it is possible to give explicit values for the optimal LOOCV parameter for RBF interpolation. We do not claim that this is a practical example – we give it only in the hope that it may provide some small insight into the difficult general problem of RBF parameter selection.

Acknowledgments. Work supported by the ex-60% funds of the University of Verona.

References

[1] L.. Bos and S. De Marchi. Univariate Radial Basis Functions with Compact Support Cardinal Functions East J. Approx., Vol. 14 (1) (2008), 69 – 80.

[2] L. Bos and U. Maier. On the asymptotics of points which maximize determinants of the form $det[g(|x_i - x_j|)]$ in "Advances in Multivariate Approximation", Mathematical Research, (W. Haussmann *et al.* Eds.), Vol. 107, pp. 107 – 128, Wiley-Vch, Berlin, 1999.

[3] L. Bos and U. Maier. On the Asymptotics of Fekete-Type Points for Univariate Radial Basis Interpolation J. of Approx. Theory **119**, 252 – 270 (2002).

[4] M. Buhmann. Radial Basis Functions Cambridge U. Press (2003).

[5] G. Fasshauer. Meshfree Approximation Methods with Matlab World Scientific Press (2007).

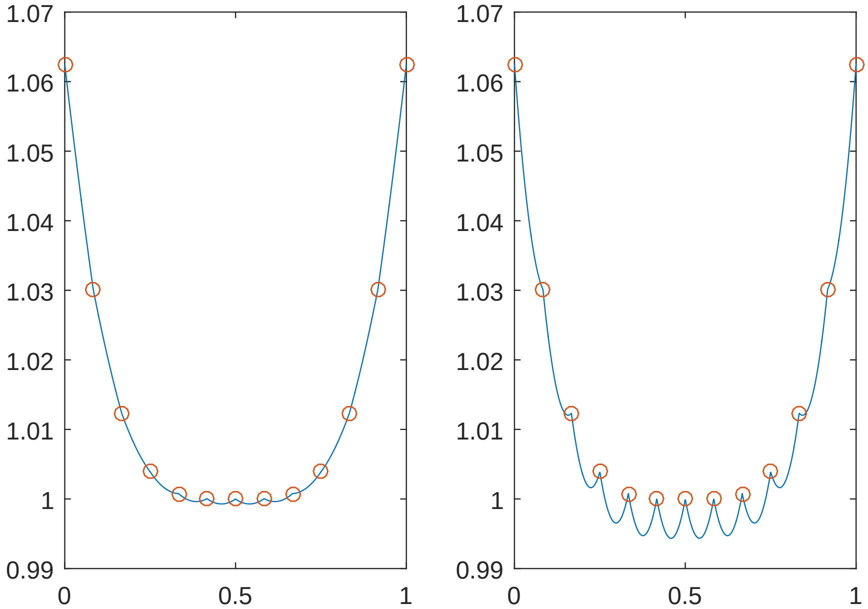

Figure 1: Left: LOOCV interpolant; Right: MLE interpolant

[6] S. Lophaven, H. B. Nielsen, and J. Sondergaard. DACE, A Matlab Kriging Toolbox Technical Report IMM-TR-2002-12, Technical Univ. of Denmark, (2002).

[7] H. Wendland. Scattered Data Approximation Cambridge U. Press (2005).