ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

J. Math. Anal. Appl. 320 (2006) 279–292

Journal of
MATHEMATICAL
ANALYSIS AND
APPLICATIONS

www.elsevier.com/locate/jmaa

Positive definite matrices and differentiable reproducing kernel inequalities

Jorge Buescu [a,*,1], A.C. Paixão [b]

[a] *Departamento Matemática, Inst. Sup. Técnico, Portugal*
[b] *Departamento Mecânica, ISEL, Portugal*

Received 8 March 2005

Available online 15 August 2005

Submitted by H.M. Srivastava

Abstract

Let $I \subseteq \mathbb{R}$ be a interval and $k : I^2 \to \mathbb{C}$ be a reproducing kernel on I. By the Moore–Aronszajn theorem, every finite matrix $k(x_i, x_j)$ is positive semidefinite. We show that, as a direct algebraic consequence, if $k(x, y)$ is appropriately differentiable it satisfies a 2-parameter family of differential inequalities of which the classical diagonal dominance is the order 0 case. An application of these inequalities to kernels of positive integral operators yields optimal Sobolev norm bounds.

Keywords: Positive definite matrices; Reproducing kernels; Inequalities

1. Introduction and definitions

Given a set E, a positive definite matrix in the sense of Moore (see, e.g., Moore [6], Aronszajn [1]) is a function $k : E \times E \to \mathbb{C}$ such that

* Corresponding author.
E-mail address: jbuescu@math.ist.utl.pt (J. Buescu).
1 The author acknowledges partial support by CAMGSD through FCT/POCTI/FEDER.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2005.06.088

$$\sum_{i,j=1}^{n} k(x_i, x_j)\overline{\xi_i}\xi_j \geqslant 0 \tag{1.1}$$

for all $n \in \mathbb{N}$ $(x_1, \dots, x_n) \in E^n$ and $(\xi_1, \dots, \xi_n) \in \mathbb{C}^n$; that is, all finite square matrices M of elements $m_{ij} = k(x_i, x_j)$, $i, j = 1, \dots, n$, are positive semidefinite.

From (1.1), it is easily shown that a positive definite matrix in the sense of Moore has the following properties: (1) it is conjugate symmetric, that is, $k(x, y) = \overline{k(y, x)}$ for all $x, y \in E$, (2) it satisfies $k(x, x) \geqslant 0$ for all $x \in E$, and (3) $|k(x, y)|^2 \leqslant k(x, x)k(y, y)$ for all $x, y \in E$. We refer to inequality (3) as the basic 'diagonal dominance' inequality; it implies the previous inequality as a particular case on the diagonal $y = x$.

The theorem of Moore–Aronszajn [1,6] provides an equivalent characterization of positive definite matrices in the sense of Moore as *reproducing kernels*. It states that $k : E \times E \to \mathbb{C}$ is a positive definite matrix in the sense of Moore if and only if there exists a (uniquely determined) Hilbert space H_k composed of functions on E such that

$$\forall y \in E, \quad k(x, y) \in H_k \text{ as a function of } x, \tag{1.2}$$

and

$$\forall x \in E,\ \forall f \in H_k, \quad f(x) = \big\langle f(y), k(y, x)\big\rangle_{H_k}. \tag{1.3}$$

Properties (1.2) and (1.3) are jointly called the *reproducing property of k in H_k*. The function k itself is called a *reproducing kernel on E* and the associated (and unique) Hilbert space H_k a *reproducing kernel Hilbert space*; see, e.g., Saitoh [8]. Throughout this paper we deal with the case where $E = I \subseteq \mathbb{R}$ is a real nonempty interval.

2. Two results on positive definite and semidefinite matrices

In this section we prove two results on finite matrices. Both admit a weak and a strong version and are stated accordingly. In the weak version, positive semidefiniteness of a matrix is assumed in the hypothesis and the conclusion is that an associated matrix is positive semidefinite. If the hypothesis is strengthened to positive definiteness, the corresponding statement is that the associated matrix is positive definite. The version relevant for reproducing kernels, to be used in Sections 3 and 4, is the semidefinite version, as should be clear from Section 1.

Let m and n be positive integers and A be a square matrix of order $n(m+1)$. In Proposition 2.1 we denote by A^{pq} the order n square submatrices of A, with $p, q = 0, \dots, m$, resulting from the partition of A into the $m+1$ square blocks defined by $[A^{pq}]_{ij} \equiv [a_{ij}^{pq}] = [a_{st}]$ for $s = i + pn$, $t = j + qn$ and $i, j = 1, \dots, n$.

Proposition 2.1. *Let A be an $n(m+1)$ square matrix. For each $\mathbf{X} = (X_0, \dots, X_m) \in \mathbb{C}^{m+1}$, define the $n \times n$ matrix*

$$\mathcal{A}(\mathbf{X}) = \sum_{p,q=0}^{m} A^{pq} X_p \overline{X_q}.$$

(a) *If A is positive semidefinite, then for any choice of* $\mathbf{X} \in \mathbb{C}^{m+1}$ *the matrix* $\mathcal{A}(\mathbf{X})$ *is positive semidefinite.*
(b) *If A is positive definite, then for any nonzero choice of* $\mathbf{X} \in \mathbb{C}^{m+1}$ *the matrix* $\mathcal{A}(\mathbf{X})$ *is positive definite.*

Proof. We begin by proving (a); the proof of (b) will differ on minor details.

Suppose A is semidefinite. Write $\mathcal{A} = [\alpha_{ij}]_{i,j=1,\dots,n}$, where $\alpha_{ij} = \sum_{p,q=0}^{m} a_{ij}^{pq} X_p \overline{X_q}$. We are required to show that, for an arbitrary choice of $(\xi_1, \dots, \xi_n) \in \mathbb{C}^n$, we have $\sum_{i,j=1}^{n} \alpha_{ij} \xi_i \overline{\xi_j} \geqslant 0$.

Since A is positive semidefinite, for any $(\zeta_1, \dots, \zeta_{n(m+1)}) \in \mathbb{C}^{n(m+1)}$ we have

$$\sum_{s,t=1}^{n(m+1)} a_{st} \zeta_s \overline{\zeta_t} \geqslant 0.$$

For $p = 0, 1, \dots, m$, $i = 1, \dots, n$ and $s = i + pn$, choose $\zeta_s = X_p \xi_i$. Then

$$\begin{aligned} \sum_{s,t=1}^{n(m+1)} a_{st} \zeta_s \overline{\zeta_t} &= \sum_{p,q=0}^{m} \sum_{s=pn+1}^{(p+1)n} \sum_{t=qn+1}^{(q+1)n} a_{st} \zeta_s \overline{\zeta_t} = \sum_{p,q=0}^{m} \sum_{i,j=1}^{n} a_{ij}^{pq} X_p \xi_i \overline{X_q} \overline{\xi_j} \\ &= \sum_{i,j=1}^{n} \left(\sum_{p,q=0}^{m} a_{pq}^{kl} X_p \overline{X_q} \right) \xi_i \overline{\xi_j} = \sum_{i,j=1}^{n} \alpha_{ij} \xi_i \overline{\xi_j} \geqslant 0. \end{aligned} \tag{2.1}$$

Therefore $\mathcal{A}(\mathbf{X})$ is positive semidefinite, as stated.

The proof of (b) is analogous, with the additional observation that since A is positive definite, $\sum_{s,t=1}^{n(m+1)} a_{st} \zeta_s \overline{\zeta_t} = 0$ holds if and only if $\zeta_s = 0$ for $s = 1, \dots, n(m+1)$. With our previous choice $\zeta_s = X_p \xi_i$ for $p = 0, 1, \dots, m$, $i = 1, \dots, n$ and $s = i + pn$, it follows from (2.1) that $\mathcal{A}(\mathbf{X}) = \sum_{i,j=1}^{n} \alpha_{ij} \xi_i \overline{\xi_j} = 0$ only if $\xi_i = 0$ for all $i = 1, \dots, n$ or $X_p = 0$ for all $p = 0, 1, \dots, m$. Since the last case is ruled out by hypothesis, it follows that $\mathcal{A}(\mathbf{X})$ is positive definite. □

Proposition 2.2. *Let T be a square matrix of order $(m+2)$ partitioned in the block form*

$$T = \left[\begin{array}{c|ccc} b & c_0 & \cdots & c_m \\ \hline d_0 & a_{00} & \cdots & a_{0m} \\ \vdots & \vdots & & \vdots \\ d_m & a_{m0} & \cdots & a_{mm} \end{array} \right]$$

and consider the square matrix A of order $m+1$ defined by

$$[A_{ij}]_{i,j=0}^{m} \equiv [a_{ij} b - d_i c_j]_{i,j=0}^{m}.$$

(a) *If T is positive semidefinite, then A is positive semidefinite.*
(b) *If T is positive definite, then A is positive definite.*

Proof. As in the previous result, we start by proving statement (a). We begin by considering the case where $b = 0$. Then the basic properties of semidefinite matrices imply that

$c_i = d_j = 0$ for all $i, j = 0, \dots, m$. Therefore in this case the matrix $[A_{ij}]_{i,j=0}^m$ is the zero matrix and the assertion is trivially verified.

From now on we assume, without loss of generality, that $b > 0$. Under this assumption we use this entry of T as a pivot and perform Gaussian elimination. Multiplying, for each $i = 0, \dots, m$, the $(i+2)$th row of T by b and subtracting from the resulting row the product of d_i by the first row we obtain the matrix T' defined by

$$T' = \left[\begin{array}{c|ccc} b & c_0 & \cdots & c_m \\ \hline 0 & A_{00} & \cdots & A_{0m} \\ \vdots & \vdots & & \vdots \\ 0 & A_{m0} & \cdots & A_{mm} \end{array}\right],$$

where $[A_{ij}]_{i,j=0}^m \equiv [a_{ij}b - d_i c_j]_{i,j=0}^m$. We now prove that this $(m+1) \times (m+1)$ matrix is positive semidefinite. To this end, it is sufficient to show that every principal minor of A is non-negative.

Let $M = \{0, 1, \dots, m\}$, $S_l = \{i_1, \dots, i_l\}$ be an arbitrary subset of l different elements of M (obviously $l \leqslant m + 1$). Let A_{S_l} be the order l principal submatrix of A associated with S_l. Consider the order $l + 1$ principal submatrix T_{S_l} of T defined by

$$T_{S_l} = \left[\begin{array}{c|c} b & C_{S_l} \\ \hline D_{S_l} & A_{S_l} \end{array}\right],$$

where C_{S_l} is the line vector $[c_j]_{j \in S_l}$ and D_{S_l} is the column vector $[d_j]_{j \in S_l}$.

It is immediate to recognize that A_{S_l} is obtained from T_{S_l} by performing the exact same procedure leading from T to A, i.e., Gaussian elimination using $b > 0$ as a pivot and consideration of the resulting principal submatrix complementary to b. Computation of the corresponding determinants then yields $|A_{S_l}| = b^{l-1}|T_{S_l}|$. Observing that T_{S_l} is a principal submatrix, and thus $|T_{S_l}|$ a principal minor, of T, which by hypothesis is positive semidefinite, we conclude that $|A_{S_l}| \geqslant 0$. Since A_{S_l} is an arbitrary principal minor of A, it follows that A is positive semidefinite.

To prove statement (b), observe that the hypothesis that T is positive definite implies that in the previous reasoning strict inequalities are valid for the minors under consideration. Since T, and therefore A, are Hermitian, this is sufficient to ensure that A is positive definite, as stated. □

3. Differential inequalities for reproducing kernels

Let $I \subseteq \mathbb{R}$ be a nontrivial interval and $k : I \times I \to \mathbb{C}$. We define the operators $\Delta_{x,h}$ and $\Delta_{y,h}$ by

$$\Delta_{x,h} k(x, y) = k(x + h, y) \quad \text{and} \quad \Delta_{y,h} k(x, y) = k(x, y + h) \tag{3.1}$$

for any $h \in \mathbb{R}$ such that $(x + h, y)$ and $(x, y + h)$ lie in I^2. In a similar way, for $m \in \mathbb{N}$ and $h \in \mathbb{R}$ denote by I_{mh} the set of all $x \in I$ such that $x + mh \in I$. For sufficiently small $|h|$, I_{mh} is nonempty. We define $\delta_{mh} : I_{mh}^2 \to \mathbb{C}$ by

$$\delta_{mh}(x,y) = (\Delta_{y,h}-1)^m(\Delta_{x,h}-1)^m k(x,y)$$
$$= \sum_{p,q=0}^{m} (-1)^{p+q}\binom{m}{p}\binom{m}{q} k(x+ph, y+qh). \tag{3.2}$$

We then have the following statement.

Lemma 3.1. *If $k(x,y)$ is a reproducing kernel on I^2, then $\delta_{mh}(x,y)$ is a reproducing kernel on I^2_{mh}.*

Proof. Let $n \in \mathbb{N}$ and $(x_1,\ldots,x_n) \in I^n_{mh}$. We are required to show that the order n square matrix

$$\mathcal{A} = [\alpha_{ij}]_{i,j=1}^n = \left[\delta_{mh}(x_i,x_j)\right]_{i,j=1}^n$$

is positive semidefinite.

Define $x_{np+i} = x_i + ph$ for every $p = 0,\ldots,m$ and consider the $n(m+1)$ square matrix $A = [a_{st}]_{s,t=1}^{n(m+1)} = [k(x_s,x_t)]_{s,t=1}^{n(m+1)}$. Since k is by hypothesis a reproducing kernel, A is positive semidefinite. We may therefore apply Proposition 2.1 to A. Writing $A^{pq} = [a^{pq}_{ij}]_{i,j=1}^n = [k(x_i+ph, x_j+qh)]_{i,j=1}^n$ for $p,q = 0,\ldots,m$, we conclude that for any choice of $(X_0,\ldots,X_m) \in \mathbb{C}^{m+1}$ the order n square matrix $\sum_{p,q=0}^m A^{pq}X_p\overline{X_q}$ is positive semidefinite. Choosing $X_p = (-1)^p\binom{m}{p}$, $p = 0,\ldots,m$, this statement implies that the order n square matrix

$$\left[\sum_{p,q=0}^{m}(-1)^{p+q}\binom{m}{p}\binom{m}{q}k(x_i+ph, x_j+qh)\right]_{i,j=1}^n = \left[\delta_{mh}(x_i,x_j)\right]_{i,j=1}^n = \mathcal{A}$$

is positive semidefinite. This finishes the proof. □

From now on, if $x \in I$ is a boundary point of I, a limit at x will mean the one-sided limit as $y \to x$ with $y \in I$ (note, however, that we do not suppose I to be closed or bounded, so that this remark need not apply).

Definition 3.2. Let $I \subset \mathbb{R}$ be an interval. A function $k : I^2 \to \mathbb{C}$ is said to *be of class $\mathcal{S}_n(I^2)$* if, for every $m_1 = 0,1,\ldots,n$ and $m_2 = 0,1,\ldots,n$, the partial derivatives $\frac{\partial^{m_1+m_2}}{\partial y^{m_2}\partial x^{m_1}}k(x,y)$ are continuous in I^2.

Remark 3.3. From Definition 3.2 it follows that $\mathcal{S}_n(I^2) \subset \mathcal{S}_m(I^2)$ if $n > m$ and $C^{2n}(I^2) \subset \mathcal{S}_n(I^2) \subset C^n(I^2)$. It is clear that a function of class $\mathcal{S}_n(I^2)$ will not in general be in $C^{n+1}(I^2)$; note, however, that in class $\mathcal{S}_n(I^2)$ the order of differentiation for all mixed partial derivatives of orders up to $2n$ is immaterial.

Lemma 3.4. *Let $I \subset \mathbb{R}^2$ be an interval and suppose $k : I^2 \to \mathbb{C}$ is of class $\mathcal{S}_1(I^2)$. Then*

$$\frac{\partial^2}{\partial y\partial x}k(x,y) = \lim_{h\to 0}\frac{1}{h^2}(\Delta_{y,h}-1)(\Delta_{x,h}-1)k(x,y).$$

Proof. Suppose first that k is real-valued. For t and h small enough (and of the correct sign if x or y are boundary points) that all relevant quantities are in I, define $\varphi(t) = k(x+t, y+h) - k(x+t, y)$. Then

$$\begin{aligned}\delta_h(x,y) &= k(x+h, y+h) - k(x+h, y) - k(x, y+h) + k(x,y)\\ &= (\Delta_{y,h} - 1)(\Delta_{x,h} - 1)k(x,y)\\ &= \varphi(h) - \varphi(0)\end{aligned} \tag{3.3}$$

with $\varphi \in C^1(I_h)$, where I_h is the compact interval of extremes 0 and h, since $\frac{\partial k}{\partial x}(x,y) \in C^1(I)$. Hence $\delta_h(x,y) = \varphi'(\theta_1)h$ with $|\theta_1| < |h|$ and

$$\lim_{h\to 0} \frac{\delta_h(x,y)}{h^2} = \lim_{h\to 0} \frac{\varphi'(\theta_1)}{h} = \lim_{h\to 0} \frac{1}{h}\left(\frac{\partial k}{\partial x}(x+\theta_1, y+h) - \frac{\partial k}{\partial x}(x+\theta_1, y)\right).$$

Since $k_1(x,y) = \frac{\partial^2}{\partial y \partial x} k(x,y) \in C(I^2)$, we have

$$\frac{\partial k}{\partial x}(x+\theta_1, y+h) - \frac{\partial k}{\partial x}(x+\theta_1, y) = k_1(x+\theta_1, y+\theta_2)h \quad \big(|\theta_2| < |h|\big),$$

and therefore

$$\lim_{h\to 0} \frac{\delta_h(x,y)}{h^2} = \lim_{h\to 0} k_1(x+\theta_1, y+\theta_2) = k_1(x,y).$$

This concludes the proof in the case where k is real-valued. If k is complex-valued, application of the above conclusions to its real and imaginary parts separately yields the lemma. □

Proposition 3.5. *Let $I \subset \mathbb{R}^2$ be an interval and suppose $k: I^2 \to \mathbb{C}$ is of class $\mathcal{S}_n(I^2)$. Then, for every $m_1, m_2 = 0, \ldots, n$,*

$$\frac{\partial^{m_1+m_2}}{\partial y^{m_2} \partial x^{m_1}} k(x,y) = \lim_{h\to 0} \frac{1}{h^{m_1+m_2}} (\Delta_{y,h} - 1)^{m_2} (\Delta_{x,h} - 1)^{m_1} k(x,y).$$

Proof. We first concentrate on the case $m_1 = m_2 \leqslant n$. For $n = 0$ the statement is empty; $n = 1$ is the content of Lemma 3.4. We proceed by induction. Suppose that, for some $m < n$,

$$k_m(x,y) = \frac{\partial^{2m}}{\partial y^m \partial x^m} k(x,y) = \lim_{h\to 0} \frac{1}{h^{2m}} (\Delta_{y,h} - 1)^m (\Delta_{x,h} - 1)^m k(x,y).$$

Clearly $k_m(x,y) = \frac{\partial^{2m}}{\partial y^m \partial x^m} k(x,y)$ satisfies the hypotheses of Lemma 3.4. Therefore

$$\begin{aligned}\frac{\partial^{2(m+1)}}{\partial y^{m+1} \partial x^{m+1}} k(x,y) &= \frac{\partial^2}{\partial y \partial x} k_m(x,y)\\ &= \lim_{h\to 0} \frac{(\Delta_{y,h} - 1)(\Delta_{x,h} - 1)k_m(x,y)}{h^2}\\ &= \lim_{h\to 0} \frac{(\Delta_{y,h} - 1)^{m+1}(\Delta_{x,h} - 1)^{m+1}k(x,y)}{h^{2(m+1)}},\end{aligned} \tag{3.4}$$

establishing the result for $m_1 + 1 = m_2 + 1 \leqslant n$.

Suppose now, without loss of generality, that $m_1 < m_2 < n$. A straightforward application of the definition of derivative yields

$$\begin{aligned}\frac{\partial^{m_1+m_2}k}{\partial y^{m_2}\partial x^{m_1}}k(x,y) &= \frac{\partial^{m_2-m_1}}{\partial y^{m_2-m_1}}k_{m_1}(x,y) \\ &= \lim_{h\to 0}\frac{(\Delta_{y,h}-1)^{m_2-m_1}k_{m_1}(x,y)}{h^{m_2-m_1}} \\ &= \lim_{h\to 0}\frac{(\Delta_{y,h}-1)^{m_2}(\Delta_{x,h}-1)^{m_1}k(x,y)}{h^{m_1+m_2}}. \end{aligned} \tag{3.5}$$

The case $m_2 < m_1 < n$ is completely analogous. This finishes the proof. $\square$

Theorem 3.6. *Let $I \subseteq \mathbb{R}$ be an interval and $k(x,y)$ be a reproducing kernel of class $\mathcal{S}_n(I^2)$. Then, for all $0 \leqslant m \leqslant n$,*

$$k_m(x,y) \equiv \frac{\partial^{2m}}{\partial y^m \partial x^m}k(x,y)$$

is a reproducing kernel of class $\mathcal{S}_{n-m}(I^2)$.

Proof. Since in the case $n = 0$ the statement is empty, we shall assume that $n \geqslant 1$. It is immediate from Definition 3.2 that, if k is in class $\mathcal{S}_n(I^2)$, then k_m is in class $\mathcal{S}_{n-m}(I^2)$ for $0 \leqslant m \leqslant n$. Then, by Proposition 3.5, we have

$$\begin{aligned} k_m(x,y) &= \lim_{h\to 0}\frac{\sum_{p,q=0}^{m}(-1)^{p+q}\binom{m}{p}\binom{m}{q}k(x+ph,y+qh)}{h^{2m}} \\ &= \lim_{h\to 0}\frac{\delta_{mh}(x,y)}{h^{2m}}, \end{aligned} \tag{3.6}$$

where the limit is to be understood as the appropriate one-sided limit if $(x,y) \in \partial(I^2)$. By Lemma 3.1, δ_{mh} is a reproducing kernel on I^2_{mh}. Then, for every positive integer l, $(x_1,\dots,x_l) \in \operatorname{int} I^l$, $(\xi_1,\dots,\xi_l) \in \mathbb{C}^l$ and sufficiently small $|h|$ we have

$$\sum_{i,j=1}^{l} k_m(x_i,x_j)\xi_i\overline{\xi_j} = \lim_{h\to 0}\frac{1}{h^{2m}}\sum_{i,j=1}^{l}\delta_{mh}(x_i,x_j)\xi_i\overline{\xi_j} \geqslant 0. \tag{3.7}$$

By continuity of k_m, inequality (3.7) holds for boundary points in I^2 (if they exist) with the interpretation of partial derivatives as appropriate one-sided limits. Thus inequality (3.7) holds for all $(x_1,\dots,x_l) \in I^l$ and every choice of l and $(\xi_1,\dots,\xi_l) \in \mathbb{C}^l$. Therefore, $k_m(x,y)$ is by the Moore–Aronszajn theorem a reproducing kernel on I^2 for each $m = 1,\dots,n$. $\square$

Corollary 3.7. *Let $I \subseteq \mathbb{R}$ be an interval and $k(x,y)$ be a reproducing kernel on I of class $\mathcal{S}_n(I^2)$. Then for all $x \in I$ and all $0 \leqslant m \leqslant n$ we have*

$$\frac{\partial^{2m}k}{\partial y^m \partial x^m}(x,x) \geqslant 0.$$

Proof. Immediate from Theorem 3.6 and the diagonal dominance inequality for reproducing kernels. □

Theorem 3.8. *Let $I \subseteq \mathbb{R}$ be an interval and $k(x, y)$ be a reproducing kernel of class $\mathcal{S}_n(I^2)$. Then, for every $m_1, m_2 = 0, 1, \ldots, n$ and all $x, y \in I$ we have*

$$\left|\frac{\partial^{m_1+m_2}}{\partial y^{m_2}\partial x^{m_1}} k(x, y)\right|^2 \leqslant \frac{\partial^{2m_1}}{\partial y^{m_1}\partial x^{m_1}} k(x, x) \frac{\partial^{2m_2}}{\partial y^{m_2}\partial x^{m_2}} k(y, y). \tag{3.8}$$

Proof. For $n = 0$ the statement yields $|k(x, y)|^2 \leqslant k(x, x)k(y, y)$, which is the classical diagonal dominance inequality for positive matrices in the sense of Moore. From here on we assume that $n \geqslant 1$.

We start by showing the particular 1-parameter case of (3.8) corresponding to $m_1 = m, m_2 = 0$. This corresponds to the statement that, for all $x, y \in I$ and every $0 \leqslant m \leqslant n$,

$$\left|\frac{\partial^m}{\partial x^m} k(x, y)\right|^2 \leqslant \frac{\partial^{2m}}{\partial y^m \partial x^m} k(x, x)k(y, y). \tag{3.9}$$

This inequality will in the end of the proof be extended to the general 2-parameter family of inequalities (3.8).

Since $k(x, y)$ is of class $\mathcal{S}_n(I^2)$, Proposition 3.5 with $m_1 = m_2 = m$ implies that

$$\frac{\partial^{2m} k}{\partial y^m \partial x^m}(x, x) = \lim_{h\to 0} \frac{1}{h^{2m}} (\Delta_{y,h} - 1)^m (\Delta_{x,h} - 1)^m k(x, x) \tag{3.10}$$

and

$$\frac{\partial^m k}{\partial x^m}(x, y) = \lim_{h\to 0} \frac{1}{h^m} (\Delta_{x,h} - 1)^m k(x, y) \tag{3.11}$$

for every $0 \leqslant m \leqslant n$.

Using (3.10) and (3.11) we may write

$$\begin{aligned} &\frac{\partial^{2m} k}{\partial y^m \partial x^m}(x, x)k(y, y) - \left|\frac{\partial^m k}{\partial x^m}(x, y)\right|^2 \\ &\quad= \lim_{h\to 0} \frac{1}{h^{2m}} \big(\big[(\Delta_{y,h} - 1)^m (\Delta_{x,h} - 1)^m k(x, x)\big]k(y, y) \\ &\qquad - \big[(\Delta_{x,h} - 1)^m k(x, y)\big]\big[\overline{(\Delta_{x,h} - 1)^m k(x, y)}\big]\big). \end{aligned} \tag{3.12}$$

We define the functions

$$\Psi(x, y, h) = \big[(\Delta_{y,h} - 1)^m (\Delta_{x,h} - 1)^m k(x, x)\big]k(y, y), \tag{3.13}$$

$$\Phi(x, y, h) = \big[(\Delta_{x,h} - 1)^m k(x, y)\big]\big[\overline{(\Delta_{x,h} - 1)^m k(x, y)}\big]. \tag{3.14}$$

We now show that $\Psi(x, y, h) - \Phi(x, y, h) \geqslant 0$ for all x, y, h where both quantities are defined. This will imply that, for all $x, y \in \operatorname{int} I$ and sufficiently small $|h|$ the finite increment on the right-hand side of (3.12) is greater or equal than zero, which in turn implies, taking the limit as $h \to 0$, that (3.12) is greater or equal than zero, establishing inequality (3.9). If x or y are boundary points of I the inequality follows from taking one-sided limits and using continuity.

Expanding (3.13) and (3.14) we obtain

$$\begin{aligned}
&\Psi(x,y,h)\\
&\quad=\left[\left(\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\Delta_{y,h}^{i}\right)\left(\sum_{j=0}^{m}(-1)^{m-j}\binom{m}{j}\Delta_{x,h}^{j}\right)k(x,x)\right]k(y,y)\\
&\quad=\left(\sum_{i,j=0}^{m}(-1)^{2m-i-j}\binom{m}{i}\binom{m}{j}\Delta_{y,h}^{i}\Delta_{x,h}^{j}k(x,x)\right)k(y,y)\\
&\quad=\left(\sum_{i,j=0}^{m}(-1)^{i+j}\binom{m}{i}\binom{m}{j}k(x+ih,x+jh)\right)k(y,y)
\end{aligned}\tag{3.15}$$

and

$$\begin{aligned}
&\Phi(x,y,h)\\
&\quad=\left(\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\Delta_{x,h}^{i}k(x,y)\right)\overline{\left(\sum_{j=0}^{m}(-1)^{m-j}\binom{m}{j}\Delta_{x,h}^{j}k(x,y)\right)}\\
&\quad=\sum_{i,j=0}^{m}(-1)^{2m-i-j}\binom{m}{i}\binom{m}{j}\Delta_{x,h}^{i}k(x,y)\overline{\Delta_{x,h}^{j}k(x,y)}\\
&\quad=\sum_{i,j=0}^{m}(-1)^{i+j}\binom{m}{i}\binom{m}{j}k(x+ih,y)\overline{k(x+jh,y)}.
\end{aligned}\tag{3.16}$$

Hence

$$\begin{aligned}
&\Psi(x,y,h)-\Phi(x,y,h)\\
&\quad=\sum_{i,j=0}^{m}(-1)^{i+j}\binom{m}{i}\binom{m}{j}\left[k(x+ih,x+jh)k(y,y)-k(x+ih,y)\overline{k(x+jh,y)}\right].
\end{aligned}\tag{3.17}$$

Using conjugate symmetry of k, we obtain

$$\begin{aligned}
&\Psi(x,y,h)-\Phi(x,y,h)\\
&\quad=\sum_{i,j=0}^{m}(-1)^{i+j}\binom{m}{i}\binom{m}{j}\left[k(x+ih,x+jh)k(y,y)-k(x+ih,y)k(y,x+jh)\right]\\
&\quad=\sum_{i,j=0}^{m}(-1)^{i+j}\binom{m}{i}\binom{m}{j}A_{ij},
\end{aligned}\tag{3.18}$$

where the A_{ij} are defined by

$$A_{ij}=k(x+ih,x+jh)k(y,y)-k(x+ih,y)k(y,x+jh)\tag{3.19}$$

for $i,j=0,\ldots,m$.

Consider the set of real numbers

$$X = \{y, x, x+h, \ldots, x+mh\} = \{y, x+ih\colon\ i = 0, \ldots, m\}.$$

Define the Hermitian $(m+2)\times(m+2)$ matrix T_X associated to X by

$$T_X = \begin{bmatrix} k(y,y) & k(y,x) & k(y,x+h) & \cdots & k(y,x+mh) \\ k(x,y) & k(x,x) & k(x,x+h) & & k(x,x+mh) \\ k(x+h,y) & k(x+h,x) & k(x+h,x+h) & \cdots & k(x+h,x+mh) \\ \vdots & & & & \vdots \\ k(x+ih,y) & \cdots & \cdots & \cdots & k(x+ih,x+mh) \\ \vdots & & & & \vdots \\ k(x+mh,y) & \cdots & \cdots & \cdots & k(x+mh,x+mh) \end{bmatrix}.$$

The fact that k is by hypothesis a positive definite matrix in the sense of Moore implies that the matrix T_X is positive semidefinite. Therefore, we may apply Proposition 2.2 to T_X. With the choices $b = k(y,y)$, $a_{ij} = k(x+ih, x+jh)$, $d_i = k(x+ih, y)$, $c_j = k(y, x+jh)$ for $i, j = 0, \ldots, m$, direct application of Proposition 2.2 implies that the matrix

$$\begin{aligned} [A_{ij}]_{i,j=0}^m &= [a_{ij}b - c_i d_j]_{i,j=0}^m \\ &= \bigl[k(x+ih, x+jh)k(y,y) - k(x_i+h, y)k(y, x+jh)\bigr]_{i,j=0}^m \end{aligned} \tag{3.20}$$

is positive semidefinite.

Finally, we show that $\Psi(x,y,h) - \Phi(x,y,h) \geqslant 0$. Define $\xi_i = (-1)^i\binom{m}{i}$ for $0 \leqslant i \leqslant m$. Since A is positive definite we have by definition

$$\sum_{i,j=0}^m A_{ij}\xi_i\overline{\xi_j} = \sum_{i,j=0}^m (-1)^{i+j}\binom{m}{i}\binom{m}{j}A_{ij} \geqslant 0.$$

Hence, by (3.17) we have $\Psi(x,y,h) - \Phi(x,y,h) \geqslant 0$.

This completes the proof of the 1-parameter family of inequalities (3.9). We next show that this implies the full 2-parameter family of inequalities (3.8).

Since k is a reproducing kernel of class $\mathcal{S}_n(I^2)$, by Theorem 3.6 k_m is a reproducing kernel of class $\mathcal{S}_{n-m}(I^2)$ for every $0 \leqslant m \leqslant n$. Let $0 \leqslant m_1 \leqslant m_2 \leqslant n$. Then $k_{m_1}(x,y) = \frac{\partial^{2m_1}}{\partial y^{m_1}\partial x^{m_1}}k(x,y)$ is a reproducing kernel of class $\mathcal{S}_{n-m_1}(I^2)$. We may write

$$\frac{\partial^{m_1+m_2}}{\partial y^{m_2}\partial x^{m_1}}k(x,y) = \frac{\partial^{m_2-m_1}}{\partial y^{m_2-m_1}}\frac{\partial^{2m_1}}{\partial y^{m_1}\partial x^{m_1}}k(x,y) = \frac{\partial^{m_2-m_1}}{\partial y^{m_2-m_1}}k_{m_1}(x,y). \tag{3.21}$$

Since $m_2 - m_1 \leqslant n - m_1$, application of (3.9) to k_{m_1} yields

$$\left|\frac{\partial^{m_2-m_1}}{\partial y^{m_2-m_1}}k_{m_1}(x,y)\right|^2 \leqslant k_{m_1}(x,x)\frac{\partial^{2(m_2-m_1)}}{\partial y^{(m_2-m_1)}\partial x^{(m_2-m_1)}}k_{m_1}(y,y).$$

Hence

$$\left|\frac{\partial^{m_2+m_1}}{\partial y^{m_2}\partial x^{m_1}}k(x,y)\right|^2 \leqslant \frac{\partial^{2m_1}}{\partial y^{m_1}\partial x^{m_1}}k(x,x)\frac{\partial^{2m_2}}{\partial y^{m_2}\partial x^{m_2}}k(y,y)$$

as stated.

Finally, it is easily shown (e.g., [3, Lemma 2.3]) that if $k(x, y)$ is a conjugate symmetric function such that $\frac{\partial^{m_1+m_2}}{\partial y^{m_2}\partial x^{m_1}}k(x, y)$ exists for all $x, y \in I$, then for all $x, y \in I$ the partial derivative $\frac{\partial^{m_1+m_2}}{\partial x^{m_2}\partial y^{m_1}}k(y, x)$ exists and satisfies

$$\frac{\partial^{m_1+m_2}}{\partial y^{m_2}\partial x^{m_1}}k(x, y) = \overline{\frac{\partial^{m_1+m_2}}{\partial x^{m_2}\partial y^{m_1}}k(y, x)}. \tag{3.22}$$

The proof of the case $0 \leqslant m_2 \leqslant m_1 \leqslant n$ in (3.8) follows immediately from the case $0 \leqslant m_1 \leqslant m_2 \leqslant n$ using (3.22). This completes the proof. □

Remark 3.9. Setting $n = 0$ in Theorem 3.8 yields the statement that if the reproducing kernel $k(x, y)$ is continuous then the basic diagonal dominance inequality $|k(x, y)|^2 \leqslant k(x, x)k(y, y)$ holds. Even though continuity is not necessary, this means that the diagonal dominance inequality for reproducing kernels may be thought of as the particular case $n = 0$ in Theorem 3.8. In this precise sense, Theorem 3.8 yields a 2-parameter family of inequalities which is the generalization of the basic diagonal dominance inequality for (adequately) differentiable reproducing kernels.

Remark 3.10. It is enlightening to observe that Theorem 3.8 and the corresponding inequalities (3.8) have a straightforward interpretation within the general theory of reproducing kernels. We next offer a brief sketch of the relevant constructions; for details on what follows see, e.g., Krein [5] or Saitoh [9].

Given a reproducing kernel $k : E \times E \to \mathbb{C}$, to each $x \in E$ we associate an abstract symbol e_x. In the complex space of finite linear combinations of such symbols we introduce an inner product defined by

$$\langle e_x, e_y\rangle \equiv k(x, y) \tag{3.23}$$

for all $x, y \in E$. Identifying vectors of zero norm yields a pre-Hilbert space, the completion of which is a Hilbert space H_k. This space is unique up to Hilbert space isomorphism and is the Moore–Aronszajn reproducing kernel Hilbert space.

Let now $E = I \subset \mathbb{R}$, and suppose that all the partial derivatives

$$\frac{\partial^{m_1+m_2}}{\partial y^{m_2}\partial x^{m_1}}k(x, y)$$

exist and are continuous for all $0 \leqslant m_1, m_2 \leqslant n$ (this is precisely requiring that k be class $\mathcal{S}_n(I^2)$; see Definition 3.2). It is then possible, although rather nontrivial, to show that the corresponding Hilbert space representatives e_x, e_y have continuous Fréchet derivatives $e_x^{(m)}, e_y^{(m)}$. Also nontrivially, for each fixed $0 \leqslant m \leqslant n$ the span of $\{e_x^{(m)}\}_{x\in E}$ is dense in H_k, thus providing a natural identification between these spaces.

Once these nontrivial facts are established, one may use the representation (3.23) to obtain

$$\left|\frac{\partial^{m_1+m_2}}{\partial y^{m_2}\partial x^{m_1}}k(x, y)\right|^2 = \left|\left\langle e_y^{(m_1)}, e_x^{(m_2)}\right\rangle_{H_k}\right|^2$$

$$\begin{aligned}
&\leqslant \left\| e_y^{(m_1)} \right\|_{H_k}^2 \left\| e_x^{(m_2)} \right\|_{H_k}^2 \\
&= \left\langle e_y^{(m_1)}, e_y^{(m_1)} \right\rangle_{H_k} \left\langle e_x^{(m_2)}, e_x^{(m_2)} \right\rangle_{H_k} \\
&= \frac{\partial^{2m_1}}{\partial y^{m_1} \partial x^{m_1}} k(y,y) \frac{\partial^{2m_2}}{\partial y^{m_2} \partial x^{m_2}} k(x,x), \qquad (3.24)
\end{aligned}$$

which is inequality (3.8). In fact, due to the reproducing property (1.2) and (1.3) of k in H_k, one may take in the previous argument the concrete representation $e_x = k(x, \cdot)$, $e_y = k(y, \cdot)$; due to uniqueness of H_k up to isomorphism there is no loss of generality in doing so.

Thus, from the abstract point of view of the general theory of reproducing kernels, the 2-parameter family of diagonal dominance inequalities (3.8) is a consequence of the Cauchy–Schwarz inequality in H_k. From what has been said it should be clear, however, that there are some advantages in the direct algebraic-analytical proof constructed in Section 3.

4. An application to integral operators

Throughout this section $I \subseteq \mathbb{R}$ will denote a closed, but not necessarily bounded, interval. A linear integral operator $K : L^2(I) \to L^2(I)$

$$K(\phi) = \int_I k(x,y)\phi(y)\,dy$$

with kernel $k(x,y) \in L^2(I^2)$ is said to be positive if

$$\int_I \int_I k(x,y)\overline{\phi(x)}\phi(y)\,dx\,dy \geqslant 0 \qquad (4.1)$$

for all $\phi \in L^2(I)$. The corresponding kernel $k(x,y)$ is an $L^2(I)$-*positive definite kernel*. A positive definite kernel is conjugate symmetric for almost all $x, y \in I$, so the associated operator K is self-adjoint and consequently all its eigenvalues are real and non-negative.

Definition 4.1. A positive definite kernel $k(x,y)$ in a closed interval $I \subseteq \mathbb{R}$ is said to be in class $\mathcal{A}_0(I)$ if

(1) it is continuous in I^2,
(2) $k(x,x) \in L^1(I)$,
(3) $k(x,x)$ is uniformly continuous in I.

Remark 4.2. If I is compact the first condition trivially implies the other two, so $\mathcal{A}_0(I)$ coincides with the continuous functions $C(I^2)$. Definition 4.1 is therefore especially meaningful in the case where I is unbounded. It may be shown [2] that, if k is a positive definite kernel in class $\mathcal{A}_0(I)$, then it satisfies Mercer's theorem [7], irrespective of whether I is bounded or unbounded. For this reason a positive definite kernel in class $\mathcal{A}_0(I)$ is sometimes called a Mercer-like kernel [3].

The following summarizes the properties of positive definite kernels relevant for this paper. If $k(x, y) \in L^2(I)$ is a positive definite kernel, then K is a Hilbert–Schmidt operator; in particular it is compact, so its eigenvalues have finite multiplicity and accumulate only at 0. The spectral expansion

$$k(x, y) = \sum_{i \geqslant 1} \lambda_i \phi_i(x) \overline{\phi_i(y)}, \tag{4.2}$$

holds, where the $\{\phi_i\}_{i \geqslant 1}$ are an $L^2(I)$-orthonormal set of eigenfunctions spanning the range of K, the $\{\lambda_i\}_{i \geqslant 1}$ are the nonzero eigenvalues of K and convergence of the series (4.2) is in $L^2(I)$. If, in addition, k is in class $\mathcal{A}_0(I)$, then $\forall x \in I$ $k(x, x) \geqslant 0$ and $\forall x, y \in I$ $|k(x, y)|^2 \leqslant k(x, x)k(y, y)$, eigenfunctions ϕ_i associated to nonzero eigenvalues are uniformly continuous on I, convergence of the series (4.2) is absolute and uniform on I, and the operator K is trace class and satisfies the trace formula $\int_I k(x, x)\, dx = \sum_{i \geqslant 1} \lambda_i$. In the case where I is compact, the last statements are the content of the classical theorem of Mercer; for proofs see, e.g., [7] for compact I and [2] for noncompact I. Finally, it is not difficult to show that positive definite kernels are reproducing kernels on I [3], so that the results of Section 3 apply.

Definition 4.3. Let $n \geqslant 1$ be an integer and $I \subseteq \mathbb{R}$. A positive definite kernel $k : I^2 \to \mathbb{C}$ is said to belong to class $\mathcal{A}_n(I)$ if $k \in \mathcal{S}_n(I)$ and

$$k(x, y), \frac{\partial^2 k}{\partial y \partial x}(x, y), \ldots, \frac{\partial^{2n} k}{\partial y^n \partial x^n}(x, y)$$

are in class $\mathcal{A}_0(I)$.

Obviously $\mathcal{A}_n(I) \subset \mathcal{A}_{n-1}(I) \subset \cdots \subset \mathcal{A}_1(I) \subset \mathcal{A}_0(I)$. A positive definite kernel in class $\mathcal{A}_n(I)$ thus possesses a delicate mix of local (differentiability class $\mathcal{S}_n(I)$) and global properties (integrability and uniform continuity of each k_m, $m = 0, \ldots, n$, along the diagonal $y = x$).

In what follows we suppose that k in class $\mathcal{A}_n(I)$. $H^n(I)$ denotes as usual the Sobolev space $W^{n,2}(I)$ normed by $\|\phi\|^2_{H^n(I)} = \sum_{m=0}^n \|\phi^{(m)}\|^2_{L^2(I)}$.

It is proved in [3] that, for k in class $\mathcal{A}_n(I)$, every eigenfunction ϕ_i associated to a nonzero eigenvalue is in $C^n(I) \cap H^n(I)$ and that, for each $m = 0, \ldots, n$ the expansion

$$k_m(x, y) = \frac{\partial^{2m} k}{\partial y^m \partial x^m}(x, y) = \sum_{i \geqslant 1} \lambda_i \phi_i^{(m)}(x) \overline{\phi_i^{(m)}(y)} \tag{4.3}$$

is uniformly and absolutely convergent in I^2. Note that for $n \geqslant 1$ this is, in general, *not* the eigenfunction series corresponding to (4.2) for k_m. In the special case where $n = 0$ and I is compact this result reduces to Mercer's theorem.

For k in class $\mathcal{A}_n(I)$, we set $\mathcal{K}_m \equiv \int_I k_m(x, x)\, dx$ for each $m = 0, \ldots, n$. From Theorem 3.8 it follows that $0 \leqslant |k_m(x, y)|^2 \leqslant k_m(x, x)k_m(y, y)$ for all $x, y \in I$. Thus for each $m = 0, \ldots, n$, $\mathcal{K}_m > 0$ unless $k_m(x, y)$ is identically zero. We define

$$C_n = \mathcal{K}_0^{1/2} \left(\sum_{m=0}^n \mathcal{K}_m \right)^{1/2}. \tag{4.4}$$

Theorem 4.4. *Let $k(x,y)$ be a positive definite kernel in class $\mathcal{A}_n(I)$, and let ϕ_i be an eigenfunction of $k(x,y)$ associated to a nonzero eigenvalue λ_i. Then ϕ_i is in $C^n(I) \cap H^n(I)$ and*

$$\|\phi_i\|_{H^n(I)} \leqslant \frac{C_n}{\lambda_i}\|\phi_i\|_{L^2(I)}, \tag{4.5}$$

where C_n is given by (4.4).

Theorem 4.4 is proved in [4] as a direct consequence of the general two-parameter inequality (3.8) of Theorem 3.8 for positive definite kernels. As also shown in [4], Theorem 4.4 is sharp: the bound it provides is optimal.

All these results—differentiability of eigenfunctions, uniform convergence of eigenseries expansion, norm bounds in Sobolev space—depend critically on positive definiteness of the kernel k and are, in general, false if k is not a positive definite kernel.

Acknowledgment

The authors acknowledge the helpful comments of an anonymous referee.

References

[1] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–404.
[2] J. Buescu, Positive integral operators in unbounded domains, J. Math. Anal. Appl. 296 (2004) 244–255.
[3] J. Buescu, A. Paixão, Positivity, differentiability and integral equations on unbounded domains, in preparation.
[4] J. Buescu, A. Paixão, Inequalities for differentiable reproducing kernels and an application to positive operators, in preparation.
[5] M.G. Krein, Hermitian positive kernels on homogeneous spaces I, Amer. Math. Soc. Transl. (2) 34 (1963) 69–108.
[6] E.H. Moore, General Analysis, Mem. Amer. Philos. Soc. Part I, 1935, Mem. Amer. Philos. Soc. Part II, 1939.
[7] F. Riesz, B. Nagy, Functional Analysis, Ungar, New York, 1952.
[8] S. Saitoh, Theory of Reproducing Kernels and its Applications, Pitman Res. Notes Math. Ser., vol. 189, Longman, 1988.
[9] S. Saitoh, Integral Transforms, Reproducing Kernels and their Applications, Pitman Res. Notes Math. Ser., vol. 369, Longman, 1997.