Korovkin Theorems and applications in approximation theory and numerical linear algebra

Stefano Serra Capizzano*

Dept. of Science and High Technology, Insubria University - Como

(stefano.serrac@uninsubria.it)

November 1, 2018

Abstract

We first review the classical Korovkin Theorem concerning the approximation in sup norm of continuous functions $C(K)$ defined over a compact domain K of $\mathbb{R}^N$, $N \geq 1$, via a sequence of linear positive operators acting from $C(K)$ into itself. Then we sketch its extension to the periodic case and, by using these general tools, we prove the first and the second Weierstrass theorem, regarding the approximation of continuous functions via Bernstein polynomials and Cesaro sums, respectively. As a second, more unexpected, application we show the use of the Korovkin theory for the fast solution of a large Toeplitz linear system $A_n x = b$, by preconditioned conjugate gradients (PCG) or PCG-NE that is PCG for normal equations. More in detail Frobenius-optimal preconditioners are chosen in some suitable matrix algebras. The resulting approximation operators are linear and positive and the use of a properly modified Korovkin Theorem is the key for proving the spectral clustering of the Frobenius-optimal preconditioned matrix-sequence. As a consequence, the resulting PCG/PCG-NE shows superlinear convergence behavior under mild additional assumptions. Three appendices and few guided exercises end this note.

1 Introduction

The first goal is the constructive approximation of continuous functions over a compact domain K of $\mathbb{R}^N$, $N \geq 1$, via functions which are simpler from the computational viewpoint. The initial choice is the polynomial one because the evaluation at a given point of a generic polynomial just implies a finite number of arithmetic operations (sums and products). The notion of "approximation"

*The present material is extracted from Notes [12] produced as part of the exam sustained by F. Didone, M. Giannuzzi, L. Paulon and concerned the series of lectures titled "Approssimazione matriciale, precondizionamento e sistemi lineari di grandi dimensioni" that the author has given during February and April 2010 in Rome, at the Department "MeMoMat", University "La Sapienza", under invitation of Prof. Daniela Giachetti, on behalf of the PhD Program in Mathematics.

is intuitive; nevertheless we spend few words for giving a formal definition. We "replace" the given object, that is our continuous function f, with a polynomial p which is close to it. Since $C(K)$ endowed with the sup-norm forms a Banach space (that is complete), it is natural to decide that the distance for measuring the closeness of two generic functions f and g belonging to $C(K)$ should be $d(f,g) = \|f-g\|_{\infty,K}$ where $\|\cdot\|_{\infty,K}$ indicates the sup-norm over K i.e.

$$\|h\|_{\infty,K} = \sup_{x\in K} |h(x)|, \quad h \in C(K).$$

Owing to the Weierstrass theorem (concerning the existence of the minimum/maximum of a continuous function defined over a compact set) and since K is a compact set of $\mathbb{R}^N$, $N \geq 1$, it is evident that the sup can be plainly replaced by max that is

$$\|h\|_{\infty,K} = \max_{x\in K} |h(x)|, \quad h \in C(K). \tag{1}$$

Here we just recall that the notion of compactness coincides with requiring that K is closed and bounded, because K is a subset of $\mathbb{R}^N$ with finite N.

At this point the problem is correctly posed. More specifically, the choice of the space of polynomials is motivated by the "computational" requirements. Is the choice also well motivated from the "approximation" point of view?

The latter question is answered by the Weierstrass theorem (concerning the polynomial approximation [19]).

Theorem 1.1 (Weierstrass)**.** *Let f be a continuous function of $C(K)$ with $K \subset \mathbb{R}^N$, $N \geq 1$, compact set. For every $\epsilon > 0$, there exists p_ϵ polynomial such that*

$$\|f - p_\epsilon\|_{\infty,K} \leq \epsilon.$$

Of course, if instead of $C(K)$, we consider the set of periodic continuous functions with period 2π in every direction

$$C_{2\pi} = \{f : \mathbb{R}^N \to \mathbb{C}, f \text{ continuous and periodic i.e. } f(x) = f(x \bmod 2\pi)\},$$

then the natural space of approximation again in the sup norm becomes that of trigonometric polynomials. Also in this case we have a theorem due to Weierstrass (see e.g. [19, 41]).

Theorem 1.2 (Weierstrass)**.** *Let f be a 2π periodic continuous function of $C_{2\pi}$ in N dimensions, $N \geq 1$. For every $\epsilon > 0$, there exists p_ϵ polynomial such that*

$$\|f - p_\epsilon\|_\infty \leq \epsilon.$$

As a tool from proving Theorems 1.1 and 1.2, we introduce the Korovkin theory in the case of continuous functions over a compact set and in the case of periodic continuous functions and we exploit the good features of the Bernstein polynomials and of the Cesaro sums. As a second, more unexpected, application of the Korovkin theory, we consider its for the fast solution of a large Toeplitz linear system $A_n x = b$, by preconditioned conjugate gradients (PCG) or PCG-NE that is PCG for normal equations. More in detail Frobenius-optimal preconditioners are chosen in some matrix algebras, associated with computationally attractive fast transforms. The resulting approximation operators are linear and positive and the use of a properly modified Korovkin Theorem is the key for proving the spectral clustering of the Frobenius-optimal preconditioned matrix-sequence. As a consequence, the resulting PCG shows superlinear convergence behavior under mild additional assumptions.

The notes are organized as follows. In section 2 we report some basic definitions and tools. In Section 3 we report the Korovkin Theorem, its proof, and some basic variants, whose analysis is completed in the last two appendices (see also the exercises in Section 10). Then, with the help of the Korovkin theory and of the Bernstein polynomials presented in Section 4.1, we produce a constructive proof of Theorem 1.1 in Section 4. As a byproduct, we furnish a tool for proving theorems of Weierstrass type in many different contexts (different spaces endowed with various norms, topologies, different models of approximation etc). As a second, more unexpected, application we show in Section 5 the use of the Korovkin theory for the fast solution of a large Toeplitz linear system $A_n x = b$, by preconditioned conjugate gradients (PCG). More in detail Frobenius-optimal preconditioners are chosen in some proper matrix algebras. The resulting approximation operators are linear and positive and the use of a properly modified Korovkin Theorem is the key for proving the spectral clustering of the Frobenius-optimal preconditioned matrix-sequence. As a consequence, the resulting PCG shows superlinear convergence behavior under mild additional assumptions. A conclusion section ends the notes (Section 6), together with three appendices (Sections 7, 8 , and 9) and a section devoted to exercises (Section 10).

2 Preliminary definitions and notions

We present the Korovkin Theorem [17, 19]. We emphasize the simplicity of the assumptions, the possibility of their plain checking, the strength of the thesis which can be adapted in a variety of settings (different spaces endowed with various norms, topologies, different models of approximation etc).

More specifically, given a sequence of operators from $C(K)$ in itself, it is sufficient to verify their linearity and positivity and their convergence to g, when applied to g, for a finite number of test functions g (which can be chosen as simple polynomials). The conclusion is the pointwise convergence to the identity over $C(K)$, that is the convergence for any continuous function. Last but not the least the proof can be made very essential with tools of elementary Calculus.

We now start with the basic definition of linear positive operator (LPO) and with definition of sequence of approximating operators (for giving a precise notion to the "pointwise convergence to the identity").

Definition 2.1. *Let $\mathcal{S}$ be a vector space of functions with values in the field $\mathbb{K}$ (with $\mathbb{K}$ being either $\mathbb{R}$ or $\mathbb{C}$) and let Φ be an operator from $\mathcal{S}$ to $\mathcal{S}$. We define the pair of properties of Φ:*

1. *for any choice of α and β in $\mathbb{K}$ and for any choice of f and g in $\mathcal{S}$, we have $\Phi(\alpha f + \beta g) = \alpha\Phi(f) + \beta\Phi(g)$ (linearity);*
2. *for any choice of $f \geq 0$, $f \in \mathcal{S}$, we have $\Phi(f) \geq 0$ (positivity).*

An operator Φ satisfying both conditions is linear and positive. We write in short LPO.

Definition 2.2. *Let* $(\mathcal{S}, \|\cdot\|)$ *be a Banach space of functions with values in the field* $\mathbb{K}$ *(with* $\mathbb{K}$ *being either* $\mathbb{R}$ *or* $\mathbb{C}$*) and let* $\{\Phi_n\}$ *be a sequence of operators each of them from* $\mathcal{S}$ *in* $\mathcal{S}$. *The sequence is a sequence of approximating operators or, more briefly an approximation process, if for every* $f \in \mathcal{S}$ *we have*

$$\lim_{n\to\infty} \|\Phi_n(f) - f\| = 0.$$

(In other words $\{\Phi_n\}$ *converges pointwise to the identity operator from* $(\mathcal{S}, \|\cdot\|)$ *into itself).*

3 Korovkin Theorem

In its simplest one-dimensional version the Korovkin Theorem says that if a sequence of LPOs (see Definition 2.1) applied to g approximates uniformly g, with g being one of the test functions 1, x, and x^2, then it approximates all continuous functions, that is the sequence represents an approximation of the identity operator, from the space of the continuous functions into itself, in the sense of Definition 2.2.

Theorem 3.1. *Let* K *be a compact subset* $\in \mathbb{R}$ *and let* $\{\Phi_n\}$ *be a sequence of linear and positive operators* $\Phi_n : C(K) \to C(K)$ *on the Banach space* $(C(K), \|\cdot\|_{\infty,K})$. *Assume that*

$$||\Phi_n(g) - g||_{\infty,K} \to 0, \quad for\ n \to \infty, \quad \forall g \in \{1, x, x^2\},$$

or $\Phi_n(g)$ *approximates the function* g *as* n *goes to infinity (this is called the Korovkin test). Then*

$$||\Phi_n(f) - f||_{\infty,K} \to 0, \quad for\ n \to \infty, \quad \forall f \in C(K),$$

or $\Phi_n(f)$ *approximates all continuous functions defined over a compact set* K, *as* n *tends to infinity.*

Proof The given assumption on the test functions

$$||\Phi_n(g) - g||_{\infty,K} \to 0, \quad for\ n \to \infty,$$

can be written as

$$\Phi_n(g(y))(x) = g(x) + \epsilon_n(g(y))(x)$$

where $\epsilon_n(g(y))(x)$ represents the error we make approximating $g(x)$ with $\Phi_n(g(y))(x)$ and hence it is such that $||\epsilon_n(g(y))||_{\infty,K} \to 0$. We want to prove that

$$||\Phi_n(f) - f||_{\infty,K} \to 0 \quad \text{for } n \to \infty,\ \forall f \in C(K)$$

and therefore we fix $f \in C(K)$ and $\epsilon > 0$, and, equivalently to the thesis, we want to prove that there exists $\bar{n}$ (as a function of f and ϵ) such that for any $n \geq \bar{n}$ and for every $x \in K$ we have $|\Phi_n(f(y))(x) - f(x)| < \epsilon$. Therefore we manipulate the latter quantity and more precisely we write $\Delta_n(f)(x) = |\Phi_n(f(y))(x) - f(x)|$ and

$$\begin{aligned} \Delta_n(f)(x) &= |\Phi_n(f(y))(x) - 1 \cdot f(x)| \\ &= |\Phi_n(f(y))(x) - (\Phi_n(1)(x) - \epsilon_n(1)(x))f(x)| \qquad (2) \end{aligned}$$

since $1 = \Phi_n(1)(x) - \epsilon_n^{(1)}(x)$.

Using the triangle inequality, from (2) and by using linearity first and then again the triangle inequality, we obtain that

$$\begin{aligned}\Delta_n(f)(x) &\leq |\Phi_n(f(y))(x) - \Phi_n(1)f(x)| + |f(x)||\epsilon_n(1)(x)| \\ &\leq |\Phi_n(f(y))(x) - \Phi_n(1)f(x)| + \|f\|_{\infty,K}\|\epsilon_n(1)\|_{\infty,K} \\ &= |\Phi_n(f(y))(x) - \Phi_n(f(x))(x)| + \|f\|_{\infty,K}\|\epsilon_n(1)\|_{\infty,K} \\ &\leq |\Phi_n(f(y) - f(x))(x)| + \|f\|_{\infty,K}\|\epsilon_n(1)\|_{\infty,K}.\end{aligned}$$

We exploit again the positivity of $\Phi_n(\cdot)$ and we find

$$\Delta_n(f)(x) \leq \Phi_n(|f(y) - f(x)|)(x) + \frac{\epsilon}{4}, \qquad \forall n \geq \bar{n}_1$$

since $\|\epsilon(1)_n(x)\|_{\infty,K} \to 0$ for n tending to ∞.

The idea of the proof is to bound from above the quantity $|f(y) - f(x)|$, uniformly with respect to $x, y \in K$ and using only test functions.

We preliminarily observe that if K is a compact set then $C(K) = UC(K)$, with $UC(K)$ denoting the set of all uniformly continuous functions. As a consequence we have

$$\forall \tilde{\epsilon} > 0 \quad \exists \delta > 0 : |x - y| < \delta \Rightarrow |f(x) - f(y)| < \tilde{\epsilon} \quad \forall x, y \in K$$

(the Cantor Theorem stating that continuity and uniform continuity are the same notion, when the domain is compact).

Thus, in general, we obtain $|f(x) - f(y)| \leq \tilde{\epsilon}$ if the points x and y are close enough. The latter condition can be rewritten as follows

$$|f(x) - f(y)| \leq \tilde{\epsilon}\chi_{\{z:|z-x|<\delta\}}(y) + 2||f||_{\infty,K}\chi_{\{z:|z-x|\geq\delta\}}(y).$$

However the righthand-side is not continuous, and it cannot be in general an argument for the operator $\Phi_n(\cdot)$. We want to majorize it by using a continuous function: we will succeed by find a bound from above made by linear combinations of the test functions. First we observe that

$$|z - x| \geq \delta \Leftrightarrow \frac{|z - x|}{\delta} \geq 1,$$

which in turn is equivalent to

$$\frac{(z - x)^2}{\delta^2} \geq 1.$$

In conclusion in the set the inequality is satisfied we can write

$$\chi_{\{z:|z-x|\geq\delta\}}(y) = 1 \leq \frac{(y - x)^2}{\delta^2}.$$

Otherwise we have

$$\chi_{\{z:|z-x|\geq\delta\}}(y) = 0 \leq \frac{(y - x)^2}{\delta^2}.$$

Therefore, uniformly with respect to $x, y \in K$, for any $\tilde{\epsilon} > 0$ we find $\delta = \delta_{\tilde{\epsilon}}\ 0$ such that

$$|f(x) - f(y)| \leq \tilde{\epsilon} + 2\|f\|_{\infty,K}\frac{(y-x)^2}{\delta^2}.$$

Hence, by applying the operator $\Phi_n(\cdot)$, for every $n \geq \bar{n}_1$ we find that

$$\begin{aligned} \Delta_n(f)(x) &\leq \frac{\epsilon}{4} + \Phi_n\left(\tilde{\epsilon} + 2\frac{\|f\|_{\infty,K}}{\delta^2}(y-x)^2\right)(x) \\ &= \frac{\epsilon}{4} + \tilde{\epsilon}\Phi_n(1)(x) + 2\frac{\|f\|_{\infty,K}}{\delta^2}\Phi_n(y^2 - 2xy + x^2)(x). \qquad (3) \end{aligned}$$

Moreover $\Phi_n(1)(x) = (1 + \epsilon_n(1)(x))$ since the constant 1 is a test function. Therefore the righthand-side in (3) can be written as

$$\frac{\epsilon}{4} + \tilde{\epsilon}(1+\epsilon_n(1)(x)) + 2\frac{\|f\|_{\infty,K}}{\delta^2}\{x^2 + \epsilon_n(y^2)(x) - 2x[x + \epsilon_n(y)(x)] + x^2(1+\epsilon_n(1)(x))\}.$$

Now we choose $\tilde{\epsilon} = \frac{\epsilon}{4}$ and we observe that, by definition of limit, there exist $\bar{n}_2, \bar{n}_3$ such that for $n \geq \bar{n}_2$ we have $\epsilon_n(1)(x) \leq 1$, uniformly with respect to $x \in K$, and for $n \geq \bar{n}_3$ we have

$$\|f\|_{\infty,K}\|\epsilon_n(y^2)(x) - 2x\epsilon_n(y)(x) + x^2\epsilon_n(1)(x)\|_{\infty,K} \leq \frac{\delta^2\tilde{\epsilon}}{2}.$$

Finally the proof is concluded by taking $n \geq \bar{n}$, with $\bar{n} = \max\{\bar{n}_1, \bar{n}_2, \bar{n}_3\}$ (depending on both f and ϵ), and by observing that

$$\Delta_n(f)(x) \leq \frac{\epsilon}{4} + 2\tilde{\epsilon} + 2\frac{\|f\|_{\infty,K}}{\delta^2}\{\epsilon_n(y^2)(x) - 2x\epsilon_n(y)(x) + x^2\epsilon_n(1)(x)\} \leq \frac{\epsilon}{4} + 3\tilde{\epsilon} = \epsilon.$$

□

In the following section, in order to prove the Weierstrass Theorem, we will propose some special polynomial sequence of linear positive operators $\{\Phi_n\}$ satisfying the Korovkin test (with $\Phi_n(f)$ being a polynomial for every $f \in C(K)$). However we should not forget that the Korovkin Theorem is very general (and flexible). Indeed no assumptions are required on the regularity of $\Phi_n(f)$. In reality, by following step by step the proof given above, we discover that proof is still valid even if $\Phi_n(f)$ is not necessarily continuous. In fact we only need that it makes sense to write $\|\Phi_n(f)\|_{\infty,K}$ or equivalently that $\Phi_n(f)$ is essentially bounded ($\Phi_n(f) \in L^\infty(K)$).

3.1 Few variations and a list of applications

As anticipated in the last section, the Korovkin Theorem is very flexible. Here we list a few changes that can be made by obtaining a rich series of new statements. Then we list a series a interesting applications

First Case: Let K be a compact subset of $\mathbb{R}^N$ and $T = \{1, x_1, x_2, \ldots, x_N, \|x\|_2^2\}$ the set of test functions. The proof of theorem is formally the same, but x must be regarded as a vector of size N. Regarding the uniform continuity the right characterization to be chosen is as follows

$$\forall\tilde{\epsilon} > 0 \quad \exists\delta > 0 : \|x - y\|_2 < \delta \Rightarrow |f(x) - f(y)| < \tilde{\epsilon} \quad \forall x, y \in K.$$

As a consequence the key majorization is the one reported below.

$$\begin{aligned} |f(y)-f(x)| &\leq \tilde{\epsilon}\chi_{\{z:||z-x||_2\leq\tilde{\delta}\}}(y)+2\|f\|_{\infty,K}\ \chi_{\{z:||z-x||_2\geq\tilde{\delta}\}}(y) \\ &\leq \tilde{\epsilon}+2\|f\|_{\infty,K}\frac{\|y-x\|_2^2}{\tilde{\delta}^2}, \end{aligned} \tag{4}$$

since $\|z-x\|_2 \geq \tilde{\delta} \Rightarrow \frac{\|z-x\|_2^2}{\tilde{\delta}^2} \geq 1$. Noting that

$$\|y-x\|_2^2 = (y^T - x^T)(y-x) = \|x\|_2^2 + ||y||_2^2 - 2\sum_{j=1}^{d} x_j y_j$$

we conclude that the function appearing in (4), with respect to the dummy variable y, is just a linear combination of test functions and therefore the rest of the proof is identical (see Appendix B).

Second Case: Let $C_{2\pi}$ be the set of 2π-periodic functions. The set of the test functions is given by

$$T = \{1, e^{\sqrt{-1}x}\}.$$

We also use the function $e^{-\sqrt{-1}x}$ but since it is the conjugate of the second test function we do not need to put it in the Korovkin test. Indeed, if f is real valued then $f = f^+ - f^-$ with $f^+ = \max\{0, f\}$, $f^- = \max\{0, -f\}$ being both nonnegative so that $\Phi(f)$ is real valued if f is and $\Phi(\cdot)$ is linear and positive. As a consequence, for a general complex valued function f we have $\Phi(\bar{f}) = \bar{\Phi(f)}$. Taking into account the latter, the proof only changes when reasoning about the uniform continuity. A possible distance between x and y is defined as

$$|e^{\sqrt{-1}x} - e^{\sqrt{-1}y}|.$$

As a consequence a clever characterization of the notion of uniform continuity is given as

$$\forall \tilde{\epsilon} > 0 \quad \exists \delta > 0 : |e^{\sqrt{-1}x} - e^{\sqrt{-1}y}| \leq \delta \Rightarrow |f(x) - f(y)| < \tilde{\epsilon} \quad \forall x, y \in \mathbb{R}.$$

Hence the key upper-bound is

$$\begin{aligned} |f(y)-f(x)| &\leq \tilde{\epsilon}\chi_{\{y:|e^{\sqrt{-1}x}-e^{\sqrt{-1}y}|\leq\delta\}}(y)+2\|f\|_\infty\ \chi_{\{y:|e^{\sqrt{-1}x}-e^{\sqrt{-1}y}|\geq\delta\}}(y) \\ &\leq \tilde{\epsilon}+2\|f\|_\infty\frac{|e^{\sqrt{-1}x}-e^{\sqrt{-1}y}|^2}{\delta^2} \\ &\leq \tilde{\epsilon}+2\frac{\|f\|_\infty}{\delta^2}(2-e^{-\sqrt{-1}x}e^{\sqrt{-1}y}-e^{\sqrt{-1}x}e^{-\sqrt{-1}y}), \end{aligned}$$

since

$$\frac{|e^{\sqrt{-1}x}-e^{\sqrt{-1}y}|^2}{\tilde{\delta}^2} \geq 1,$$

where $2 - e^{-\sqrt{-1}x}e^{\sqrt{-1}y} - e^{\sqrt{-1}x}e^{-\sqrt{-1}y}$ is the linear combination of test functions, with regard to the dummy variable y (see Appendix C, also for the multivariate setting).

Third case: With reference to Theorem 3.1, the Korovkin set $T = \{1, x, x^2\}$ can be replaced by any Chebyshev set $C = \{p_0(x), p_1(x), p_2(x)\}$ of continuous functions over the compact set K such that the 3×3 Vandermonde-like matrix

$$(p_j(x_k))_{j,k=0,1,2}$$

is invertible for every choice of points $\{x_k\}$ such that $x_0 < x_1 < x_2$, $x_0, x_1, x_2 \in K$. It is worth noticing that an analogous result does not hold in more than one dimension, due to topological reasons (see Exercise 12).

As already stressed in the introduction, among classical applications of the Korovkin theory, we can list the first and the second Weierstrass Theorems (a huge variety of applications in approximation theory can be found in the work by Altomare and coauthors [1]). As a more exotic application we finally mention the use of the Korovkin tools in the approximation of Toeplitz matrix sequences via (computationally simpler) matrix algebras [26], including the set of circulant matrices associated to the celebrated fast Fourier transform (FFT). The corresponding Korovkin test is successfully verified for all ω-circulants with $|\omega| = 1$, with all known sine, cosine, and Hartley transform spaces (see [10, 11, 2]) and is extended to the case of block structures [27] and to various operator theory settings [28, 18]: to the matrix version setting we devoted the analysis in Section 5.

4 Proof of the Weierstrass Theorem

In this section we first introduce the Bernsterin polynomials (and the related sequence of operators) and we prove linearity, positivity and the proof of the positive answer to the Korovkin test. The we give a (semi)-constructive proof of the Weierstrass Theorem. The only non-constructive tool concerns the use of the Tietze extension Theorem (see [22]).

4.1 The Bernstein polynomials

We furnish a concrete example of a sequence of linear positive operators. We will perform the Korovkin test. For the sake of simplicity we first consider the one-dimensional case with $N = 1$ and $K = [0, 1]$. The Bernstein polynomials of f, indicated by $B_n(f)$, are defined as

$$\Phi_n(f)(x) \equiv B_n(f)(x) = \sum_{\nu=0}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} f\left(\frac{\nu}{n}\right). \tag{5}$$

It is straightforward to verify that it is a linear and positive operator, since it is expressed as linear combination of nonnegative polynomials $x^\nu(1-x)^{n-\nu}$ of degree n, via the coefficients

$$\begin{pmatrix} n \\ \nu \end{pmatrix} f\left(\frac{\nu}{n}\right),$$

which are all nonnegative if $f \geq 0$. From these remarks we immediately get both positivity and linearity of the operator B_n.

4.1.1 Bernstein polynomials and Korovkin test

We now perform the Korovkin test. For $g \in \{1, x, x^2\}$ we prove that $B_n(g)$ converges uniformly to g on $[0,1]$.

For $g = 1$ we have

$$\begin{aligned} B_n(1)(x) &= \sum_{\nu=0}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} \\ &= (x + (1-x))^n = 1 \end{aligned}$$

and thus the error $\epsilon_n(1)$ is identically zero.

For $g(x) = x$ we have

$$\begin{aligned} B_n(g(y))(x) &= \sum_{\nu=0}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} \frac{\nu}{n} \\ &= \sum_{\nu=1}^{n} \frac{n!}{\nu!(n-\nu)!} \frac{\nu}{n} x^\nu (1-x)^{n-\nu} \\ &= \sum_{\nu=1}^{n} \begin{pmatrix} n-1 \\ \nu-1 \end{pmatrix} x^\nu (1-x)^{n-\nu} \\ &= \sum_{q \equiv \nu-1=0}^{n-1} \begin{pmatrix} n-1 \\ q \end{pmatrix} x^{q+1} (1-x)^{n-1-q} \\ &= x B_{n-1}(1)(x) = x \end{aligned}$$

and hence, also in this case, the error $\epsilon_n(y)$ is identically zero.

Finally we consider $g(x) = x^2$ so that

$$\begin{aligned} B_n(g(y))(x) &= \sum_{\nu=0}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} \frac{\nu^2}{n^2} \\ &= \sum_{\nu=0}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} \frac{\nu(\nu-1)}{n^2} \\ &\quad + \sum_{\nu=0}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} \frac{\nu}{n^2} \\ &= \sum_{\nu=2}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} \frac{\nu(\nu-1)}{n^2} + \frac{1}{n} B_n(y)(x) \\ &= \frac{n-1}{n} \sum_{\nu=2}^{n} \begin{pmatrix} n \\ \nu \end{pmatrix} x^\nu (1-x)^{n-\nu} \frac{\nu(\nu-1)}{n(n-1)} + \frac{1}{n} x \\ &= \frac{n-1}{n} \sum_{\nu=2}^{n} \begin{pmatrix} n-2 \\ \nu-2 \end{pmatrix} x^\nu (1-x)^{n-\nu} + \frac{1}{n} x \\ &= \frac{n-1}{n} \sum_{q \equiv \nu-2=0}^{n-2} \begin{pmatrix} n-2 \\ q \end{pmatrix} x^{q+2} (1-x)^{n-2-q} + \frac{1}{n} x \\ &= \frac{n-1}{n} x^2 B_{n-2}(1)(x) + \frac{1}{n} x = \frac{n-1}{n} x^2 + \frac{1}{n} x \\ &= x^2 + \frac{x(1-x)}{n}. \end{aligned}$$

In conclusion the error $\epsilon_n(y^2)(x) = \frac{x(1-x)}{n}$ converges uniformly to zero so that the test of Korovkin has a positive answer. Then, by invoking the Korovkin Theorem, we deduce that $\{B_n\}$ is an approximation process on $C([0,1], \|\cdot\|_\infty)$.

Remark 4.1. *We note that the error is identically zero for the first and second test functions. If the error in the third test function would have been identically zero, then we would have concluded*

$$B_n(f(y))(x) = f(x), \qquad \forall f, x, n.$$

Hence $\mathbb{C}(K)$ would coincide with space of polynomials: the latter is clearly not true, so we necessarily expected a non-zero error for $j = 2$ (refer to Exercise 2 and also Exercise 3 for a deeper understanding on the results).

4.1.2 The case of the multidimensional Bernstein polynomials

We generalize the notion of Bernstein polynomial in the multidimensional setting. We consider $K = [0,1]^N$ and on this set we define the Bernstein polynomial of f denoted by $B_n(f)$, with $n = (n_1, \ldots, n_N)$ multi-index:

$$\begin{aligned} B_n(f(y))(x) \quad = \quad & \sum_{\nu_1=0}^{n_1} \cdots \sum_{\nu_N=0}^{n_N} \binom{n_1}{\nu_1} \cdots \binom{n_N}{\nu_N} \\ & x^{\nu_1}(1-x)^{n_1-\nu_1} \cdots x^{\nu_N}(1-x)^{n_N-\nu_N} f\left(\frac{\nu_1}{n_1}, \ldots, \frac{\nu_N}{n_N}\right). \end{aligned} \tag{6}$$

It is trivial to verify that the given operator is linear and positive since it is a linear combination of nonnegative polynomials of degree n

$$x^{\nu_1}(1-x)^{n_1-\nu_1} \cdots x^{\nu_N}(1-x)^{n_N-\nu_N}$$

via the coefficients

$$\binom{n_1}{\nu_1} \cdots \binom{n_N}{\nu_N} f\left(\frac{\nu_1}{n_1}, \ldots, \frac{\nu_N}{n_N}\right),$$

which are all nonnegative if $f \geq 0$. From these observations both linearity and positivity directly follow.

4.2 A Korovkin based proof of the Weierstrass Theorem

We prove Theorem 1.1 of which we recall once again the statement and then we proceed with its proof based on the Korovkin theory.

Let f be any function belonging to $C(K)$ with $K \subset \mathbb{R}^N$, $N \geq 1$, compact set. For every $\epsilon > 0$ there exists p_ϵ polynomial such that

$$\|f - p_\epsilon\|_{\infty,K} \leq \epsilon.$$

Proof The proof is organized by steps and more precisely by treating the case of a domain K of increasing "complexity".
Step 1: $K = [0,1]$

For $K = [0,1]$ the Bernstein polynomials represent a sequence of linear and positive operators satisfying the Korovkin test. Therefore the thesis follows Theorem 3.1 that is

$$\lim_{n\to\infty} \|f(x) - B_n(f)(x)\|_{\infty,K} = 0.$$

Step 2: $K = [a,b]$
We consider $g \in C([0,1])$ such that

$$g(t) = f(a + (b-a)t).$$

Therefore, by the previous item, there exists $B_{\bar{n}_\epsilon}(g)$ satisfying the relation

$$\|\Delta_n(g)\|_{\infty,[0,1]} = \|g(t) - B_{\bar{n}_\epsilon}(g)(t)\|_{\infty,[0,1]} \le \epsilon.$$

Now we perform the change of variable $x = a + (b-a)t$ so that $t = \frac{x-a}{b-a}$ and $g\left(\frac{x-a}{b-a}\right) = f(x)$. Thus

$$\begin{aligned} \|\Delta_n(g)\|_{\infty,[0,1]} &= \left\| g\left(\frac{x-a}{b-a}\right) - B_{\bar{n}_\epsilon}(g)\left(\frac{x-a}{b-a}\right)\right\|_{\infty,[a,b]} \\ &= \left\| f(x) - B_{\bar{n}_\epsilon}(g)\left(\frac{x-a}{b-a}\right)\right\|_{\infty,[a,b]} \le \epsilon. \end{aligned}$$

Since

$$B_{\bar{n}_\epsilon}(g)\left(\frac{x-a}{b-a}\right)$$

is still a polynomial in the variable $x \in [a,b]$, the desired thesis follows.
Step 3: $K = [0,1]^N$
As observed in Section 4.1.2, the N dimensional Bernstein polynomials are linear positive operators (it plainly follows from the definition). Moreover they satisfy the Korovkin test in Theorem 8.1 (it is much easier than expected: follow Exercises 6 and 7 at the end of the note). Therefore the claimed thesis follows by Theorem 8.1 as in **Step 1**.
Step 4: $K = \prod_{i=1}^{N}[a_i, b_i]$
The proof goes along the same lines as in **Step 2**: we use N decoupled affine transformations sending $[0,1]^N$ onto K and viceversa (i.e. $x_i = a_i + (b_i - a_i)t_i$, $i = 1, \ldots, N$) and the thesis in **Step 3**.
Step 5: K **compact of** $\mathbb{R}^N$
Every compact K of $\mathbb{R}^N$ is contained in a proper N dimensional rectangle of the form $\tilde{K} = \prod_{i=1}^{N}[a_i, b_i]$. By the Tietze extension Theorem (see [22]), every function f defined on K has a continuous extension $\tilde{f}$ defined on $\tilde{K}$ (that is $\tilde{f}$ is globally continuous on $\tilde{K}$ and such that $\tilde{f}(x) = f(x)$, for every $x \in K$). Therefore, by reasoning as in **Step 4** for the function $\tilde{f}$, it is sufficient for concluding the proof. •

As it happens e.g. in the polynomial interpolation, many issues regarding the convergence with polynomial approximations strongly depend from the topology of the domain and co-domain of the continuous functions to be approximated. For instance if K is a compact set of $\mathbb{C}$, the Weierstrass Theorem is simply false if the interior part of K is non empty: the reason relies on the fact that the uniform limit over any compact set contained in a given open domain Ω of a sequence of holomorphic functions over Ω (e.g. a sequence of polynomials with Ω being the internal part of K) is still holomorphic over Ω.

4.3 An historical note: the original proof given by Weierstrass

In this section we sketch the original proof provided by Weierstrass in his work. Of course the proof is not explicitly based on the theory of linear positive operators (the Korovkin theory is more recent and is dated around 1958-1960), but it uses a special sequence of linear positive operators that is those of Gauss-Weierstrass.

The n-th Gauss-Weiertrass operator is defined as

$$(\mathcal{GW}_n(f))(x) = \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-n(t-x)^2}{2}} f(t)dt. \tag{7}$$

Unlike the Bernstein case, we observe that $(\mathcal{GW}_n(f))(x)$ is not a polynomial in general but it is a smooth function belonging to C^∞ on the whole real line $\mathbb{R}$ (thanks to the role of a Gaussian weight).

The remaining part of the proof follows according to the ideas described below:

Step 1. We take the Taylor polynomial $p_{n,m}(x)$ of degree m associated with $(\mathcal{GW}_n(f))(x)$ and it is possible to prove that

$$\lim_{m\to\infty} \|(\mathcal{GW}_n(f))(x) - p_{n,m}(x)\|_{\infty,K} = 0$$

for every compact K of $\mathbb{R}$.

Step 2. It is possible to prove that $(\mathcal{GW}_n(f))(x)$ converges uniformly to $f(x)$ for every compact K of $\mathbb{R}$, as n tends to infinity.

The two considered steps imply the Weierstass Theorem for every compact K of $\mathbb{R}$. It should be mentioned that the proof is only partially constructive since $p_{n,m}$ is defined using information on f not available in general (except via numerical evaluations).

5 Application of the Korovkin theory to preconditioning

This section deals with some applications of Korovkin theory to the preconditioning, since we are interested to design fast iterative solvers when the size n of our linear system is large. More in detail we consider the case of Toeplitz matrices and we show that the approximation of such matrices in specific matrix algebras can be reduced, at least asymptotically, to a single test on the shift matrix: the idea behind this surprising simplification is the use of the Korovkin machinery on the spectrum of the matrices approximating the Toeplitz structures.

The section is organized into three parts: in the first we give preliminary definitions and tools, in the second we present Toeplitz matrices, the space of approximation, and the approximation strategy and finally in the third part we state and prove the Korokvin theorem in the setting of Toeplitz matrix sequences with continuous symbol.

5.1 Definitions and tools

Definition 5.1. *A matrix sequence $\{A_n\}$, A_n square matrix of size n, is* clustered at $s \in \mathbb{C}$ *(in the eigenvalue sense), if for any $\varepsilon > 0$ the number of the eigenvalues of A_n off the disk*

$$D(s,\varepsilon) := \{z : |z - s| < \varepsilon\}$$

is $o(n)$. In other words

$$q_\varepsilon(n,s) := \#\{\lambda_j(A_n) : \lambda_j \notin D(s,\varepsilon)\} = o(n), \quad n \to \infty.$$

If every A_n has only real eigenvalues (at least for all n large enough), then s is real and the disk $D(s,\varepsilon)$ reduces to the interval $(s-\varepsilon, s+\varepsilon)$. The cluster is strong if the term $o(n)$ is replace by $O(1)$ so that the number of outlying eigenvalues is bounded by a constant not depending on the size n of the matrix.

We say that a preconditioner P_n is optimal for A_n if the sequence $\{P_n^{-1}A_n\}$ is clustered at one (or to any positive constant) in the strong sense. Since

$$P_n^{-1}A_n = I_n + P_n^{-1}(A_n - P_n),$$

it is evident that for the optimality we need to prove the strong cluster at zero of $\{A_n - P_n\}$; for details on this issue refer to [12]; in addition several useful results on preconditioning can be found in [3, 20], while in [15, 24] the reader can find a rich account on general Krylov methods. The (weak or general) clustering is also of interest, as a heuristic indication that the preconditioner is effective.

Remark 5.2. *The previous Definition 5.1 can be stated also for the singular values instead of the eigenvalues just replacing λ with σ and "eigenvalue" with "singular value".*

Given a matrix A we define the Schatten p norm, $p \in [1,\infty)$, as the p norm of the vector of its singular values i.e. $\|A\|_{S,p} = \left[\sum_{j=1}^n \sigma_j^p(A)\right]^{1/p}$. By taking a limit on p the Schatten ∞ norm is exactly the spectral norm i.e. the maximal singular value. Among these norms the one which interesting both from a computational viewpoint and from a theoretical viewpoint is the Schatten 2 norm which is called the Frobenius norm in the Numerical Analysis community. We define now the Frobenius norm as it is usually defined in a computational setting

$$\|A\|_F = \left[\sum_{j,k=1}^n |A_{j,k}|^2\right]^{1/2}. \tag{8}$$

Of course if A is replaced by QA with Q unitary matrix, then every column will maintain the same Euclidean length and therefore $\|QA\|_F = \|A\|_F$; on the other hand, if A is replaced by AP with P unitary matrix, then every row will maintain the same Euclidean length and therefore $\|AP\|_F = \|A\|_F$. These two statements show that $\|\cdot\|_F$ is a unitarily invariant (u.i.) norm so that $\|QAP\|_F = \|A\|_F$ for every matrix A and for every unitary matrices P and Q; a rich and elegant account on the properties of such norms can be found in [5]. Therefore, taking into account the singular value decomposition [14] of A and choosing P and Q as the transpose conjugate of the left and right vectors matrices, we conclude that $\|A\|_F = \|\Sigma\|_F$ with Σ being the diagonal matrix with the singular values $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_n(A)$. Therefore

$$\|\cdot\|_F = \|\cdot\|_{S,2},$$

so that the Schatten 2 norm has a very nice computational expression that only depends on the entries of the matrix. Furthermore the definition in (8) reveals that $\|A\|_F^2 = \langle A, A\rangle$, with $\langle A, B\rangle = \mathrm{trace}(B^*A)$ being the (positive) Frobenius scalar product. Therefore the Frobenius norm comes from a positive scalar product and consequently the space of the matrices of size n with the Schatten 2 norm is a Hilbert space: this further property represents a second important ingredient for the strategy of approximation that we follow in order to obtain a good preconditioner.

The following lemma links the notion of clustering with a quantitative estimate of the Schatten p norms.

Lemma 5.3. *Assume that a sequence of matrices $\{X_n\}$ is given and that $\|X_n\|_{S,p} = O(1)$ for some $p \in [1, \infty)$. Then $\{X_n\}$ is clustered at zero in the strong sense.*

Proof The assumption implies that there exist a constant M such that $\|X_n\|_{S,p} \leq M$ for n large enough. Consequently the proof reduces to a series of inequalities as shown below

$$\begin{aligned}
M^p &\geq \|X_n\|_{S,p}^p \quad (\text{by the assumption } \|X_n\|_{S,p} \leq M, \quad n \geq \bar{n}) \\
&= \sum_{j=1}^{n} \sigma_j^p(X_n) \quad (\text{by definition of } \|\cdot\|_{S,p}) \\
&\geq \sum_{j:\sigma_j(X_n)>\epsilon} \sigma_j^p(X_n) \quad (\text{for any fixed } \ \epsilon > 0) \\
&\geq \sum_{j:\sigma_j(X_n)>\epsilon} \epsilon^p \\
&= \epsilon^p \#\{j : \sigma_j(X_n) > \epsilon\}.
\end{aligned}$$

Therefore, for every $\epsilon > 0$, we have proved that the cardinality of the set of singular values of X_n exceeding ϵ, for n large enough, is bounded by the constant $(M/\epsilon)^p$ which, by definition, is equivalent to write that $\{X_n\}$ is clustered at zero in the strong sense. •

The result above is nicely complemented by the following remark.

Remark 5.4. *With reference to Lemma 5.3, if the assumption that $\|X_n\|_{S,p} = O(1)$ for some $p \in [1, \infty)$ is replaced by the weaker request that $\|X_n\|_{S,p} = o(n)$ for some $p \in [1, \infty)$, then the same proof of the lemma shows that $\{X_n\}$ is clustered at zero in the weak sense.*

5.2 Toeplitz sequences, approximating spaces, and approximation strategy

Given a sequence $\{U_n\}$ of unitary matrices with U_n of size n, we first define the vector spaces $\{\mathcal{A}(U_n)\}$ with

$$\mathcal{A}(U_n) = \{X = U_n D U_n^* : D \text{ is a diagonal (complex) matrix}\}.$$

By definition each $\mathcal{A}(U_n)$ is a complex vector space of dimension n with the structure of algebra, that is closed under multiplication and a fortiori, by the Cayley-Hamilton Theorem [14], under inversion. It is worth noticing that, in practice, Numerical Analysts are interested in the case where U_n represent some discrete transform with linear complexity, as the Wavelet transform, or with quasi-linear complexity (e.g. $O(n\log(n))$ arithmetic operations), as in the case of all Fourier-like, trigonometric, Hartley transforms. Now let us define the operator $\mathcal{P}_{U_n}(\cdot)$ acting on $\mathcal{M}_n(\mathbb{C})$ and taking values in $\mathcal{A}(U_n)$ where both the Hilbert spaces are equipped with the Frobenius norm $\|X\|_F^2 = \sum_{i,j} |x_{i,j}|^2$, associated with the scalar product $\langle A, B\rangle = \mathrm{trace}(B^*A)$. Then

$$\mathcal{P}_{U_n}(A) = \arg\min_{X\in\mathcal{A}(U_n)} \|A - X\|_F,$$

where the minimum exists and is unique since $\mathcal{A}(U_n)$ is a linear finite dimensional space and hence it is closed.

Lemma 5.5. *With $A, B \in \mathcal{M}_n(\mathbb{C})$ and the previous definition of $\mathcal{P}_{U_n}(\cdot)$, we have*

1. $\mathcal{P}_{U_n}(A) = U_n\sigma(U_n^*AU_n)U_n^*$, *with $\sigma(X)$ the diagonal matrix having $(X)_{i,i}$ as diagonal elements,*
2. $\mathcal{P}_{U_n}(A^*) = (\mathcal{P}_{U_n}(A))^*$,
3. $\mathrm{trace}(\mathcal{P}_{U_n}(A)) = \mathrm{trace}(A)$,
4. $\mathcal{P}_{U_n}(\cdot)$ *is linear that is* $\mathcal{P}_{U_n}(\alpha A + \beta B) = \alpha\mathcal{P}_{U_n}(A) + \beta\mathcal{P}_{U_n}(B)$ *with* $\alpha, \beta \in \mathbb{C}$, *and positive that is $\mathcal{P}_{U_n}(A)$ is Hermitian positive semi-definite if A is Hermitian positive semi-definite,*
5. $\|A - \mathcal{P}_{U_n}(A)\|_F^2 = \|A\|_F^2 - \|\mathcal{P}_{U_n}(A)\|_F^2$,
6. $|||\mathcal{P}_{U_n}(\cdot)||| = 1$ *with $|||\cdot|||$ being any dual u.i. norm.*

Proof The proof of Item 1 is a direct consequence of the minimization process, taking into account that the Frobenius norm is u.i. In reality

$$\begin{aligned}
\mathcal{P}_{U_n}(A) &= \arg\min_{\{X\in\mathcal{A}(U_n)\}} ||A - X||_F \\
&= \arg\min_{\{D \text{ diagonal matrix}\}} ||A - U_nDU_n^*||_F \\
&= \arg\min_{\{D \text{ diagonal matrix}\}} ||U_n^*AU_n - D||_F.
\end{aligned}$$

Therefore, the optimal condition is obtained by choosing $D = D_{\mathrm{Opt}}$ as the diagonal part of $U_n^*AU_n$. Now Item 2 is a direct consequence of Item 1 as well as Item 3, where we just observe that A and $U_n^*AU_n$ share the same spectrum by similarity (U_n^* is the inverse of U_n). The first part of Item 4 is again a direct consequence of the representation formula in Item 1, while for the second part we just observe that the positive semi-definiteness of A is equivalent to that of $U_n^*AU_n$ so that its diagonal $\sigma(U_n^*AU_n)$ is made up by nonnegative numbers which are the eigenvalues of the Hermitian matrix $\mathcal{P}_{U_n}(A)$. Item 5 is nothing but the Pythagora Law holding in any Hilbert space. The proof of Item 6 is reduced to prove $|||\mathcal{P}_{U_n}(A)||| \leq |||A|||$ for every complex matrix, since $\mathcal{P}_{U_n}(I) = I$. The general result is given in [30], thanks to a general variational principle. For the subsequent purposes we need in fact only the statement for the spectral norm. In that case, using the singular value decomposition [14], it is easy to see that

$$\|A\| = \max_{\|u\|_2=\|v\|_2=1} |v^*Au|, \tag{9}$$

so that

$$\begin{aligned} \|A\| &= \max_{\|u\|_2=\|v\|_2=1} |v^*Au| \\ &\geq \max_{i=1,\dots,n} |u_i^*Au_i| \\ &= \|\mathcal{P}_{U_n}(A)\| \end{aligned}$$

with $u_1, \dots, u_n$ being the orthonormal columns forming U_n and where in the last equality we used the key representation formula in Item 1. We finally observe that $\|\mathcal{P}_{U_n}\|_F = 1$ is a direct consequence of the Pythagora Law in Item 5. •

We will apply the former projection technique to Toeplitz matrix-sequences $\{A_n\}$ where the n-th matrix A_n is of the form

$$(A_n(f))_{j,k} = a_{j-k}(f) \tag{10}$$

with a_j, $j \in \mathbb{Z}$, given coefficients. In the following we are interested to the case where, given $f \in L^1(0, 2\pi)$, the entries of the matrix A_n are defined via the Fourier coefficients of f that is

$$a_j = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-\sqrt{-1}jt}\, dt.$$

The latter defines a sequence $\{T_n(\cdot)\}$ of operators, $T_n : L^1(0, 2\pi) \to \mathcal{M}_n(\mathbb{C})$, which are clearly linear, due to the linearity of the Fourier coefficients, and positive as proved in detail below. In fact, if f is real valued then it is immediate to see that the conjugate of a_j is exactly a_{-j} so that by (10) we deduce that $T_n^*(f) = T_n(f)$. Now we observe that for any pair of vectors u and v (with entries indexed from 0 to $n-1$) we have the identity

$$\begin{aligned} v^*T_n(f)u &= \sum_{j,k=0}^{n-1} \bar{v}_j (T_n(f))_{j,k} u_k \\ &= \sum_{j,k=0}^{n-1} \bar{v}_j a_{j-k}(f) u_k \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) \sum_{j,k=0}^{n-1} \bar{v}_j u_k e^{-\sqrt{-1}jt} e^{\sqrt{-1}kt}\, dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) \bar{p}_v(t) p_u(t)\, dt, \end{aligned} \tag{11}$$

where, for a given vector x of size with entries indexed from 0 to $n-1$, the symbol $p_x(t)$ stands for the polynomial

$$p_x(t) = \sum_{k=0}^{n-1} x_k e^{\sqrt{-1}kt}.$$

It is worth noticing that the map from $\mathbb{C}^n$ with Euclidean norm to the space $\mathbb{P}_{n-1}$ of polynomials of degree at most $n-1$ on the unit circle with the Haar $L^2(0, 2\pi)$ norm is an isometry, since

$$\|x\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |p_x(t)|^2 \, dt. \tag{12}$$

Now the positivity of the operator is a plain consequence of the integral representation in (11), since for every $u \in \mathbb{C}^n$ and every nonnegative $f \in L^1(0, 2\pi)$, we have

$$u^* T_n(f) u = \frac{1}{2\pi} \int_0^{2\pi} f(t) |p_u(t)|^2 \, dt \geq 0.$$

In fact, a little more effort (see [25]) leads to conclude $u^* T_n(f) u = 0$ if and only the nonnegative symbol f is zero a.e.

Therefore the composition of $\mathcal{P}_{U_n}(\cdot)$ and $T_n(\cdot)$ is also a linear and positive operator from $L^1(0, 2\pi)$ to $\mathcal{A}(U_n) \subset \mathcal{M}_n(\mathbb{C})$.

Now starting from the eigenvalues of $\mathcal{P}_{U_n}(T_n(f))$ we may construct a linear positive operator from $L^1(0, 2\pi)$ to $C_{2\pi}$; in fact we will be interested mainly to the case where $f \in C_{2\pi}$ so that we will consider the restriction of $\mathcal{P}_{U_n}(T_n(\cdot))$ to $C_{2\pi}$.

The construction proceeds as follows. Let f be a function belonging to $C_{2\pi}$. We consider the j-th eigenvalue of $\mathcal{P}_{U_n}(T_n(f))$ which is

$$\lambda_j(f, n) = (U_n^* T_n(f) U_n)_{jj}, \quad j = 0, \ldots, n-1, \tag{13}$$

which can be seen a linear positive functional with respect to the argument f that is $\lambda_j(\cdot, n) : C_{2\pi} \to \mathbb{C}$ is a linear positive functional. We now define a global operator $\Phi_n(\cdot)$ from $C_{2\pi}$ into itself as reported below

$$\Phi_n(f(y))(x) = \begin{cases} \lambda_j(f(y), n), & \text{if } x \bmod 2\pi = \frac{2\pi j}{n} \\ \text{linear interpolation at the endpoints}, & \text{if } x \bmod 2\pi \in I_{j,n} \end{cases}$$

for $j = 0, 1, ..., n-1$, $I_{j,n} = \left[\frac{2\pi j}{n}, \frac{2\pi(j+1)}{n}\right)$, and with $\lambda_n(f, n)$ set equal to $\lambda_0(f, n) = (U_n^* T_n(f) U_n)_{00}$ for imposing periodicity. We observe that for every $f \in C_{2\pi}$ the function $\Phi_n(f(t))(x)$ is also continuous and 2π periodic. Indeed, the operator

$$\Phi_n : C_{2\pi} \to C_{2\pi}$$

is linear, since the sampling points in (13) are linear functionals and the interpolation retains the property of linearity. Furthermore if f is nonnegative, then $\lambda_j(f, n)$ are all nonnegative (due the positivity of all the functionals $\lambda_j(\cdot, n)$) and the linear interpolation preserves the non-negativity. As a conclusion, $\{\Phi_n\}$ is a sequence of linear positive operators from $C_{2\pi}$ into $C_{2\pi}$ and hence we can apply the Korovkin Theorem in its periodic version.

5.3 The Korovkin Theorem in a Toeplitz setting

As a final tool we need a quantitative version of the Korovkin Theorem for trigonometric polynomials. We report it in the next lemma. Then we specialize such a result for dealing with the operators $\{\Phi_n\}$ constructed from the eigenvalue functionals in (13) and, finally, we prove the Korovkin Theorem in the Toeplitz setting.

Lemma 5.6. *Let us consider the standard Banach space $C_{2\pi}$ endowed the sup-norm. Let us denote by $T = \{1, e^{\sqrt{-1}x}\}$ the standard Korovkin set of 2π periodic test functions and let us take a sequence $\{\Phi_n\}$ of linear positive operators from $C_{2\pi}$ in itself. If for any $g \in T$*

$$\max_{g\in T} \|\Phi_n(g) - g\|_\infty = O(\theta_n),$$

with θ_n tending to zero as n tends to infinity, then, for any trigonometric polynomial p of fixed degree (independent of n), we find

$$\|\Phi_n(p) - p\|_\infty = O(\theta_n).$$

Proof We first observe that, by the linearity and by the positivity of Φ_n (see the discussion at the beginning of the first item in Section 3.1), we get

$$\|\Phi_n(e^{\pm\sqrt{-1}y})(x) - e^{\pm\sqrt{-1}x}\|_\infty = O(\theta_n).$$

Moreover, from the linearity of the space of the trigonometric polynomials, the claimed thesis is proved if we prove the statement for the monomials $e^{\pm k\sqrt{-1}x}$ for any positive fixed integer k. Therefore setting $w = e^{\sqrt{-1}y}$ and $z = e^{\sqrt{-1}x}$ we have

$$\begin{aligned}
\Phi_n(w^k)(x) - z^k &= \Phi_n(w^k)(x) - \Phi_n(1)(x)z^k + O(\theta_n) \\
&= \Phi_n(w^k)(x) - \Phi_n(z^k)(x) + O(\theta_n) \\
&= \Phi_n(w^k - z^k)(x) + O(\theta_n).
\end{aligned}$$

Now it is useful to manipulate the difference $w^k - z^k$. Actually the following simple relations hold:

$$\begin{aligned}
w^k - z^k &= (w-z)kz^{k-1} + (w-z)\cdot \\
&\quad (w^{k-1} - z^{k-1} + z(w^{k-2} - z^{k-2}) + \ldots + z^{k-2}(w-z)) \\
&= (w-z)kz^{k-1} + |w-z|^2 \tfrac{1}{(\bar{w}-\bar{z})}\cdot \\
&\quad (w^{k-1} - z^{k-1} + z(w^{k-2} - z^{k-2}) + \ldots + z^{k-2}(w-z)).
\end{aligned}$$

Furthermore, setting

$$R(w,z) = \frac{1}{(\bar{w}-\bar{z})}(w^{k-1} - z^{k-1} + z(w^{k-2} - z^{k-2}) + \ldots + z^{k-2}(w-z)),$$

we find that

$$\begin{aligned}
\|R(w,z)\|_\infty &= \|\tfrac{1}{(\bar{w}-\bar{z})}\cdot(w^{k-1} - z^{k-1} + \\
&\quad + z(w^{k-2} - z^{k-2}) + \ldots + z^{k-2}(w-z))\|_\infty \\
&= \|\tfrac{1}{(w-z)}\cdot(w^{k-1} - z^{k-1} + \\
&\quad + z(w^{k-2} - z^{k-2}) + \ldots + z^{k-2}(w-z))\|_\infty \\
&\le \textstyle\sum_{j=1}^{k-1} j = (k-1)k/2.
\end{aligned}$$

Now we exploit the linearity and the positivity of the operators $\Phi_n(\cdot)$ and the fact that z and x are constants with regard to the operator $\Phi_n(\cdot)$.

$$\begin{array}{rcl}|\Phi_n(w^k - z^k)(x)| & = & |\Phi_n((w-z)kz^{k-1})(x) + \Phi_n(|w-z|^2 R(w,z))(x)| \\ & = & |\Phi_n(w-z)(x)kz^{k-1} + \Phi_n(|w-z|^2 R(w,z))(x)| \\ & = & |O(\theta_n)kz^{k-1} + \Phi_n(|w-z|^2 R(w,z))(x)|.\end{array}$$

By linearity we have $\Phi_n(|w-z|^2 R(w,z))(x) = \Phi_n(|w-z|^2 \mathrm{Re}(R(w,z)))(x) + \sqrt{-1}\Phi_n(|w-z|^2 \mathrm{Im}(R(w,z)))(x)$ where $\mathrm{Re}(s)$ and $\mathrm{Im}(s)$ denote the real and the imaginary part of s respectively. By the triangle inequality and by the positivity of Φ_n we deduce

$$\begin{array}{rcl}|\Phi_n(|w-z|^2 R(w,z))(x)| & \leq & |\Phi_n(|w-z|^2 \mathrm{Re}(R(w,z)))(x)| + \\ & & +|\Phi_n(|w-z|^2 \mathrm{Im}(R(w,z)))(x)| \\ & \leq & \Phi_n(|w-z|^2 |\mathrm{Re}(R(w,z))|)(x) + \\ & & +\Phi_n(|w-z|^2 |\mathrm{Im}(R(w,z))|)(x) \\ & \leq & 2\Phi_n(|w-z|^2 \|R(w,z)\|_\infty)(x).\end{array}$$

Therefore, by taking into account the preceding relationships we obtain

$$\begin{array}{rcl}|\Phi_n(w^k - z^k)(x)| & \leq & |O(\theta_n)k| + \Phi_n(|w-z|^2)2\|R(w,z)\|_\infty \\ & \leq & O(\theta_n) + \Phi_n(|w-z|^2)2\|R(w,z)\|_\infty \\ & \leq & O(\theta_n) + \Phi_n(|w-z|^2)(k-1)k \\ & \leq & O(\theta_n) + \Phi_n(2 - w\bar{z} - z\bar{w})(k-1)k \\ & \leq & O(\theta_n) + (2 + O(\theta_n)) - (z + O(\theta_n))\bar{z} - \\ & & -z(\bar{z} + O(\theta_n)))(k-1)k = O(\theta_n).\end{array}$$

The latter, together with the initial equations, implies that

$$\|\Phi_n(w^k)(x) - z^k\|_\infty = O(\theta_n)$$

and the lemma is proved. •

Here we report a specific version of the above lemma, adapted to the specific sequence $\{\Phi_n\}$ constructed in the previous subsection, which is more suited for the subsequent clustering analysis.

Lemma 5.7. *Let us consider the standard Banach space $C_{2\pi}$ endowed the sup-norm, let us define $x_j^{(n)} = \frac{2\pi j}{n}$, $j = 0, \ldots, n-1$, and let us take the sequence $\{\Phi_n\}$ of linear positive operators from $C_{2\pi}$ in itself, defined via the linear eigenvalue functionals in (13). If*

$$\max_{j=0,\ldots,n-1} \|\lambda_j(e^{\sqrt{-1}x}, n) - e^{\sqrt{-1}x_j^{(n)}}\|_\infty = O(\theta_n),$$

with $\theta_n = O(n^{-1})$, for n going to infinity, then, for any trigonometric polynomial p of fixed degree (independent of n), we find

$$\|\Phi_n(p) - p\|_\infty = O(n^{-1}).$$

Proof First of all we observe that the special nature of the operators $\{\Phi_n\}$ implies that only the single test function $g(x) = e^{\sqrt{-1}x}$ must be considered.

Indeed the identity matrix I belongs to the algebra $\mathcal{A}(U_n)$ since U_n is unitary. Therefore we have no error: in fact we have

$$g = 1 \Rightarrow T_n(1) = I \Rightarrow \mathcal{P}_{U_n}(T_n(1)) = \mathcal{P}(I) = I,$$

so that $\lambda_j(1, n) = 1$ and therefore

$$\Phi_n(1)(x) \equiv 1.$$

As in the previous lemma, since $\mathcal{P}_{U_n}(A^*) = \mathcal{P}^*_{U_n}(A)$, we have $T_n(e^{-\sqrt{-1}t}) = T_n^*(e^{\sqrt{-1}t})$, $\mathcal{P}_{U_n}(T_n(e^{-\sqrt{-1}t})) = \mathcal{P}^*_{U_n}(T_n(e^{\sqrt{-1}t}))$, so that

$$\lambda_j(e^{-\sqrt{-1}t}, n) = \bar{\lambda}_j(e^{\sqrt{-1}t}, n)$$

that is in general

$$\Phi_n(\bar{f}) = \bar{\Phi}_n(f).$$

The proof now is reduced to the application of Lemma 5.6, if we prove that the given assumption

$$\max_{j=0,\ldots,n-1} \|\lambda_j(e^{\sqrt{-1}x}, n) - e^{\sqrt{-1}x_j^{(n)}}\|_\infty = O(\theta_n)$$

implies

$$\|\Phi_n(e^{\sqrt{-1}y})(x) - e^{\sqrt{-1}x}\|_\infty = O(n^{-1})$$

uniformly with respect to $x \in \mathbb{R}$. The proof is plain. In fact let $h(x) = e^{\sqrt{-1}x}$. Then $q(x) = \Phi_n(e^{\sqrt{-1}y})(x)$ is continuous, 2π periodic, piecewise linear, and such that $|h(x) - q(x)| = O(\theta_n)$ uniformly for $x = x_j^{(n)}$. We have to prove the same estimate for $x \in (x_j^{(n)}, x_{j+1}^{(n)})$. We have

$$\begin{aligned} h(x) - q(x) &= h(x) - h(x_j^{(n)}) + h(x_j^{(n)}) - q(x_j^{(n)}) + q(x_j^{(n)}) - q(x) \\ &= \alpha_1 + \alpha_2 + \alpha_3, \end{aligned}$$

where

$$\begin{aligned} \alpha_1 &= h(x) - h(x_j^{(n)}), \\ \alpha_2 &= h(x_j^{(n)}) - q(x_j^{(n)}), \\ \alpha_3 &= q(x_j^{(n)}) - q(x). \end{aligned}$$

Now $|\alpha_2| = O(\theta_n)$ by hypothesis, $|\alpha_3| = O(n^{-1})$ because q is a piecewise linear function, almost interpolating with error $O(\theta_n)$ the Lipschitz continuous function h, where the latter observation or a direct Taylor expansion implies $|\alpha_1| = O(n^{-1})$. •

Theorem 5.8. *Let us consider the standard Banach space $C_{2\pi}$ endowed the sup-norm, let us define $x_j^{(n)} = \frac{2\pi j}{n}$, $j = 0, \ldots, n-1$, and let us take the sequence $\{\Phi_n\}$ of linear positive operators from $C_{2\pi}$ in itself, defined via the linear eigenvalue functionals in (13). If*

$$\max_{j=0,\ldots,n-1} \|\lambda_j(e^{\sqrt{-1}x}, n) - e^{\sqrt{-1}x_j^{(n)}}\|_\infty = O(\theta_n),$$

with θ_n tending to zero, as n tends to infinity, then for every $f \in C_{2\pi}$ the sequence $\{T_n(f) - \mathcal{P}_{U_n}(T_n(f))\}$ is weakly clustered at zero.

Proof By definition of clustering (see Definition 5.1) and taking into consideration the singular value decomposition, see [14], given $f \in C_{2\pi}$, the thesis amounts in proving that, for every $\epsilon > 0$, there exist sequences $\{N_{n,\epsilon}\}$ and $\{R_{n,\epsilon}\}$ such that

$$T_n(f) - \mathcal{P}_{U_n}(T_n(f)) = N_{n,\epsilon} + R_{n,\epsilon}, \tag{14}$$

with $\|N_{n,\epsilon}\| < \epsilon$ and $\mathrm{rank}(R_{n,\epsilon}) \leq r(\epsilon)n$ with

$$\lim_{\epsilon \to 0} r(\epsilon) = 0.$$

We take a generic function $f \in C_{2\pi}$. By virtue of the second Weierstrass theorem (i.e. Theorem 1.2 with $N = 1$), for every $\delta > 0$ there exists p_δ for which

$$\|f - p_\delta\|_\infty < \delta$$

where p_δ is a trigonometric polynomial. Hence, taking into account the linearity of $\mathcal{P}_{U_n}(\cdot)$ and $T_n(\cdot)$ proved in Section 5.2, we obtain

$$\begin{aligned} T_n(f) - \mathcal{P}_{U_n}(T_n(f)) &= T_n(f) - T_n(p_\delta) + T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta)) \\ &\quad + \mathcal{P}_{U_n}(T_n(p_\delta)) - \mathcal{P}_{U_n}(T_n(f)) \\ &= T_n(f - p_\delta) + T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta)) + \mathcal{P}_{U_n}(T_n(p_\delta - f)) \\ &= E_1 + E_2 + E_3, \end{aligned}$$

where $E_1 = T_n(f - p_\delta)$, $E_2 = T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta))$, $E_3 = \mathcal{P}_{U_n}(T_n(p_\delta - f))$ are the error terms. At this point we observe that relations (9) and (11) implies

$$\|E_1\| \leq \|f - p_\delta\|_\infty < \delta,$$

while the previous inequality and Item 6 in Lemma 5.5, with the choice of the spectral norm, lead to

$$\|E_3\| \leq \|E_1\| < \delta.$$

Now we must prove handle the term E_2 and for it we consider the Frobenius norm, having in mind both Remark 5.4 and Item 5 in Lemma 5.5 (the Pythagora equality). In fact

$$\|E_2\|_F^2 = \|T_n(p_\delta)\|_F^2 - \|\mathcal{P}_{U_n}(T_n(p_\delta))\|_F^2$$

with

$$p(t) = p_\delta(t) = \sum_{j=-q}^{q} a_j e^{\sqrt{-1}jt}.$$

By looking at the entries of the Toeplitz matrix $T_n(p)$ generated by the polynomial $p = p_\delta$, we have $(n - 2q)\|p\|_{L^2}^2 \leq \|T_n(p)\|_F^2 \leq n\|p\|_{L^2}^2$ since

$$\sum_{j=-q}^{q} |a_j|^2 = \|p\|_{L^2}^2.$$

Therefore

$$\|T_n(p)\|_F^2 = n||p||_{L^2}^2 + O(1)$$

so that

$$\|E_2\|_F^2 = n\|p\|_{L^2}^2 + O(1) - \|\mathcal{P}_{U_n}(T_n(p))\|_F^2. \tag{15}$$

For the latter term we invoke Lemma 5.7, since

$$\begin{aligned}\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 &= \sum_{j=0}^{n-1} \sigma_j^2(\mathcal{P}_{U_n}(T_n(p))) \\ &= \sum_{j=0}^{n-1} |\lambda_j(p,n)|^2.\end{aligned}$$

Indeed, owing to the assumption

$$\max_{0\le j\le n-1} |\lambda_j(e^{\sqrt{-1}t}, n) - e^{\sqrt{-1}\frac{2\pi j}{n}}| = O(\theta_n),$$

Lemma 5.7 leads to write

$$\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 = \sum_{j=0}^{n-1} \left| p\left(\frac{2\pi j}{n} + f_{j,n}\right)\right|^2$$

with

$$|f_{j,n}| \le C\theta_n$$

and C being a constant independent both of n and j. Therefore

$$\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 = \sum_{j=0}^{n-1} \left[\left| p\left(\frac{2\pi j}{n}\right)\right|^2 + \tilde{f}_{j,n}\right]$$

which is equal to

$$= \sum_{j=0}^{n-1} \left| p\left(\frac{2\pi j}{n}\right)\right|^2 + O(n\theta_n)$$

with

$$|\tilde{f}_{j,n}| \le \tilde{C}\theta_n,$$

$\tilde{C}$ being a constant independent both of n and j, and $n\theta_n = o(n)$. The latter displayed equality is a Riemann sum of a smooth function so that

$$\begin{aligned}\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 &= \frac{n}{2\pi}\left(\frac{2\pi}{n}\sum_{j=0}^{n-1}\left|p\left(\frac{2\pi j}{n}\right)\right|^2\right) + o(n) \\ &= \frac{n}{2\pi}\left[\int_0^{2\pi} |p|^2 dt + O(n^{-1})\right] + o(n) \\ &= \frac{n}{2\pi}\int_0^{2\pi} |p|^2 dt + O(1) + o(n) \\ &= n\|p\|_{L^2}^2 + o(n)\end{aligned}$$

that is $\|E_2\|_F^2 = o(n)$, owing to (15). Now we apply the result stated in Remark 5.4 with reference to the Frobenius norm and again the singular values decomposition. Therefore, for every $\epsilon > 0$, we find sequences $\{\tilde{N}_{n,\epsilon}\}$ and $\{\tilde{R}_{n,\epsilon}\}$ such that

$$E_2 = \tilde{N}_{n,\epsilon} + \tilde{R}_{n,\epsilon}$$

$\|\tilde{N}_{n,\epsilon}\| < \epsilon/3$ and $\mathrm{rank}(\tilde{R}_{n,\epsilon}) \leq \tilde{r}(\epsilon)n$ with

$$\lim_{\epsilon\to 0} \tilde{r}(\epsilon) = 0.$$

Finally we choose $\delta = \epsilon/3$ and the proof is concluded with reference to (14) and with the equalities

$$R_{n,\epsilon} = \tilde{R}_{n,\epsilon}, \quad N_{n,\epsilon} = \tilde{N}_{n,\epsilon} + E_1 + E_3.$$

•

We follow the same proof given in Theorem 5.8. In particular, in all the relevant equations the terms $o(n)$ are replaced by terms of constant order. In the following we give all the proof for the sake of completeness. (Notice that in [26], the proof is given under a more general assumption that the grid points considered in Lemma 5.7 enjoy a property of quasi-uniform distribution, when the weak clustering of $\{T_n(f) - \mathcal{P}_{U_n}(T_n(f))\}$ is derived, or a property of uniform distribution, when the strong clustering of $\{T_n(f) - \mathcal{P}_{U_n}(T_n(f))\}$ is obtained.)

Theorem 5.9. *Let us consider the standard Banach space $C_{2\pi}$ endowed the sup-norm, let us define $x_j^{(n)} = \frac{2\pi j}{n}$, $j = 0, \dots, n-1$, and let us take the sequence $\{\Phi_n\}$ of linear positive operators from $C_{2\pi}$ in itself, defined via the linear eigenvalue functionals in (13). If*

$$\max_{j=0,\dots,n-1} \|\lambda_j(e^{\sqrt{-1}x}, n) - e^{\sqrt{-1}x_j^{(n)}}\|_\infty = O(\theta_n),$$

with $\theta_n = O(n^{-1})$, for n going to infinity, then for every $f \in C_{2\pi}$ the sequence $\{T_n(f) - \mathcal{P}_{U_n}(T_n(f))\}$ is clustered at zero.

Proof In analogy with the previous theorem, by definition of strong clustering and taking into consideration the singular value decomposition, given $f \in C_{2\pi}$, the thesis amounts in proving that, for every $\epsilon > 0$, there exist sequences $\{N_{n,\epsilon}\}$ and $\{R_{n,\epsilon}\}$ such that

$$T_n(f) - \mathcal{P}_{U_n}(T_n(f)) = N_{n,\epsilon} + R_{n,\epsilon}, \tag{16}$$

with $\|N_{n,\epsilon}\| < \epsilon$ and $\mathrm{rank}(R_{n,\epsilon}) \leq r(\epsilon)$.

We take a generic function $f \in C_{2\pi}$. By virtue of the second Weierstrass theorem (i.e. Theorem 1.2 with $N = 1$), for every $\delta > 0$ there exists p_δ for which

$$\|f - p_\delta\|_\infty < \delta$$

where p_δ is a trigonometric polynomial. Hence, taking into account the linearity of $\mathcal{P}_{U_n}(\cdot)$ and $T_n(\cdot)$ proved in Section 5.2, we obtain

$$\begin{aligned} T_n(f) - \mathcal{P}_{U_n}(T_n(f)) &= T_n(f) - T_n(p_\delta) + T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta)) \\ &\quad + \mathcal{P}_{U_n}(T_n(p_\delta)) - \mathcal{P}_{U_n}(T_n(f)) \\ &= T_n(f - p_\delta) + T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta)) + \mathcal{P}_{U_n}(T_n(p_\delta - f)) \\ &= E_1 + E_2 + E_3, \end{aligned}$$

where $E_1 = T_n(f - p_\delta)$, $E_2 = T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta))$, $E_3 = \mathcal{P}_{U_n}(T_n(p_\delta - f))$ are the error terms. At this point, we observe that relations (9) and (11) implies

$$\|E_1\| \leq \|f - p_\delta\|_\infty < \delta,$$

while the previous inequality and Item 6 in Lemma 5.5, with the choice of the spectral norm, imply

$$\|E_3\| \leq \|E_1\| < \delta.$$

Now we must prove handle the term E_2 and for it we consider the Frobenius norm, having in mind both Lemma 5.3 and Item 5 in Lemma 5.5 (the Pythagora equality). In fact

$$\|E_2\|_F^2 = \|T_n(p_\delta)\|_F^2 - \|\mathcal{P}_{U_n}(T_n(p_\delta))\|_F^2$$

with

$$p(t) = p_\delta(t) = \sum_{j=-q}^{q} a_j e^{\sqrt{-1}jt}.$$

Now looking at the entries of the Toeplitz matrix $T_n(p)$ generated by the polynomial $p = p_\delta$, we have $(n-2q)\|p\|_{L^2}^2 \leq \|T_n(p)\|_F^2 \leq n\|p\|_{L^2}^2$ since

$$\sum_{j=-q}^{q} |a_j|^2 = \|p\|_{L^2}^2.$$

Therefore

$$\|T_n(p)\|_F^2 = n||p||_{L^2}^2 + O(1)$$

so that

$$\|E_2\|_F^2 = n\|p\|_{L^2}^2 + O(1) - \|\mathcal{P}_{U_n}(T_n(p))\|_F^2. \tag{17}$$

For the latter term we invoke Lemma 5.7, since

$$\begin{aligned}
\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 &= \sum_{j=0}^{n-1} \sigma_j^2(\mathcal{P}_{U_n}(T_n(p))) \\
&= \sum_{j=0}^{n-1} |\lambda_j(p,n)|^2.
\end{aligned}$$

Indeed, owing to the assumption

$$\max_{0\leq j\leq n-1} |\lambda_j(e^{\sqrt{-1}t}, n) - e^{\sqrt{-1}\frac{2\pi j}{n}}| = O(n^{-1}),$$

Lemma 5.7 leads to deduce

$$\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 = \sum_{j=0}^{n-1} \left| p\left(\frac{2\pi j}{n} + f_{j,n}\right)\right|^2$$

with

$$|f_{j,n}| \leq \frac{C}{n}$$

and C being a constant independent both of n and j. Therefore

$$\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 = \sum_{j=0}^{n-1} \left[\left| p\left(\frac{2\pi j}{n}\right)\right|^2 + \tilde{f}_{j,n}\right]$$

which is equal to

$$= \sum_{j=0}^{n-1} \left| p\left(\frac{2\pi j}{n}\right) \right|^2 + O(1)$$

with

$$|\tilde{f}_{j,n}| \le \frac{\tilde{C}}{n}$$

and $\tilde{C}$ being a constant independent both of n and j. The latter displayed equality is a Riemann sum of a smooth function and hence

$$\begin{aligned} \|\mathcal{P}_{U_n}(T_n(p))\|_F^2 &= \frac{n}{2\pi}\left(\frac{2\pi}{n}\sum_{j=0}^{n-1}\left|p\left(\frac{2\pi j}{n}\right)\right|^2\right) + O(1) \\ &= \frac{n}{2\pi}\left[\int_0^{2\pi} |p|^2 dt + O(n^{-1})\right] + O(1) \\ &= \frac{n}{2\pi}\int_0^{2\pi} |p|^2 dt + O(1) \\ &= n\|p\|_{L^2}^2 + O(1) \end{aligned}$$

that is $\|E_2\|_F^2 = O(1)$, thanks to relation (17). Furthermore the application of Lemma 5.3 with the Frobenius norm, in connection with the use of the singular value decomposition, implies that for every $\epsilon > 0$, we find sequences $\{\tilde{N}_{n,\epsilon}\}$ and $\{\tilde{R}_{n,\epsilon}\}$ such that

$$E_2 = \tilde{N}_{n,\epsilon} + \tilde{R}_{n,\epsilon}$$

$\|\tilde{N}_{n,\epsilon}\| < \epsilon/3$ and $\mathrm{rank}(\tilde{R}_{n,\epsilon}) \le \tilde{r}(\epsilon)$. Finally we choose $\delta = \epsilon/3$ and the proof is concluded with reference to (16) and with the equalities

$$R_{n,\epsilon} = \tilde{R}_{n,\epsilon}, \quad N_{n,\epsilon} = \tilde{N}_{n,\epsilon} + E_1 + E_3.$$

•

Theorem 5.10. *Let us consider the standard Banach space $C_{2\pi}$ endowed the sup-norm, let us define $x_j^{(n)} = \frac{2\pi j}{n}$, $j = 0, \ldots, n-1$, and let us take the sequence $\{\Phi_n\}$ of linear positive operators from $C_{2\pi}$ in itself, defined via the linear eigenvalue functionals in (13). If*

$$\max_{j=0,\ldots,n-1} \|\lambda_j(e^{\sqrt{-1}x}, n) - e^{\sqrt{-1}x_j^{(n)}}\|_\infty = O(\theta_n),$$

with θ_n tending to zero, as n tends to infinity, then for every $f \in L^1(0, 2\pi)$ the sequence $\{T_n(f) - \mathcal{P}_{U_n}(T_n(f))\}$ is weakly clustered at zero.

Proof By definition of clustering (see Definition 5.1) and taking into consideration the singular value decomposition, see [14], given $f \in C_{2\pi}$, the thesis amounts in proving that, for every $\epsilon > 0$, there exist sequences $\{N_{n,\epsilon}\}$ and $\{R_{n,\epsilon}\}$ such that

$$T_n(f) - \mathcal{P}_{U_n}(T_n(f)) = N_{n,\epsilon} + R_{n,\epsilon}, \tag{18}$$

with $\|N_{n,\epsilon}\| < \epsilon$ and $\mathrm{rank}(R_{n,\epsilon}) \le r(\epsilon)n$ with

$$\lim_{\epsilon\to 0} r(\epsilon) = 0.$$

We take a generic function $f \in L^1(0, 2\pi)$. By virtue of the density of trigonometric polynomials in the space $L^1(0, 2\pi)$, in the L^1 topology, for every $\delta > 0$ there exists p_δ for which

$$\|f - p_\delta\|_{L^1} < \delta$$

where p_δ is a trigonometric polynomial. Hence, taking into account the linearity of $\mathcal{P}_{U_n}(\cdot)$ and $T_n(\cdot)$ proved in Section 5.2, we obtain

$$\begin{aligned} T_n(f) - \mathcal{P}_{U_n}(T_n(f)) &= T_n(f) - T_n(p_\delta) + T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta)) \\ &\quad + \mathcal{P}_{U_n}(T_n(p_\delta)) - \mathcal{P}_{U_n}(T_n(f)) \\ &= T_n(f - p_\delta) + T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta)) + \mathcal{P}_{U_n}(T_n(p_\delta - f)) \\ &= E_1 + E_2 + E_3, \end{aligned}$$

where $E_1 = T_n(f - p_\delta)$, $E_2 = T_n(p_\delta) - \mathcal{P}_{U_n}(T_n(p_\delta))$, $E_3 = \mathcal{P}_{U_n}(T_n(p_\delta - f))$ are the error terms. At this point we invoke a variational characterization in [30] which implies the inequality

$$\|E_1\|_{S,1} \le \|f - p_\delta\|_\infty < \delta n,$$

while the previous inequality and Item 6 in Lemma 5.5, with the choice of the Schatten 1-norm, lead to conclude

$$\|E_3\|_{S,1} \le \|E_1\|_{S,1} < \delta n.$$

Therefore, by applying result stated in Remark 5.4 with the Schatten 1-norm and by recalling the singular value decomposition, we find sequences $\{\hat{N}_{n,\delta}\}$ and $\{\hat{R}_{n,\delta}\}$ such that

$$E_1 + E_3 = \hat{N}_{n,\delta} + \hat{R}_{n,\delta}$$

$\|\hat{N}_{n,\delta}\| < \sqrt{\delta}$ and $\mathrm{rank}(\hat{R}_{n,\epsilon}) \le \sqrt{\delta} n$.

Now we must prove handle the term E_2 and for it we consider the Frobenius norm, having in mind both Remark 5.4 and Item 5 in Lemma 5.5 (the Pythagora equality). In fact

$$\|E_2\|_F^2 = \|T_n(p_\delta)\|_F^2 - \|\mathcal{P}_{U_n}(T_n(p_\delta))\|_F^2$$

with

$$p(t) = p_\delta(t) = \sum_{j=-q}^{q} a_j e^{\sqrt{-1} j t}.$$

By looking at the entries of the Toeplitz matrix $T_n(p)$ generated by the polynomial $p = p_\delta$, we have $(n - 2q)\|p\|_{L^2}^2 \le \|T_n(p)\|_F^2 \le n\|p\|_{L^2}^2$ since

$$\sum_{j=-q}^{q} |a_j|^2 = \|p\|_{L^2}^2.$$

Therefore

$$\|T_n(p)\|_F^2 = n||p||_{L^2}^2 + O(1)$$

and hence

$$\|E_2\|_F^2 = n\|p\|_{L^2}^2 + O(1) - \|\mathcal{P}_{U_n}(T_n(p))\|_F^2. \tag{19}$$

For the latter term we invoke Lemma 5.7, since

$$\begin{aligned}\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 &= \sum_{j=0}^{n-1} \sigma_j^2(\mathcal{P}_{U_n}(T_n(p))) \\ &= \sum_{j=0}^{n-1} |\lambda_j(p,n)|^2.\end{aligned}$$

Indeed, owing to the assumption

$$\max_{0\le j\le n-1} |\lambda_j(e^{\sqrt{-1}t}, n) - e^{\sqrt{-1}\frac{2\pi j}{n}}| = O(\theta_n),$$

Lemma 5.7 implies

$$\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 = \sum_{j=0}^{n-1} \left| p\left(\frac{2\pi j}{n} + f_{j,n}\right)\right|^2$$

with

$$|f_{j,n}| \le C\theta_n$$

and C being a constant independent both of n and j. Therefore

$$\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 = \sum_{j=0}^{n-1} \left[\left| p\left(\frac{2\pi j}{n}\right)\right|^2 + \tilde{f}_{j,n}\right]$$

which is equal to

$$= \sum_{j=0}^{n-1} \left| p\left(\frac{2\pi j}{n}\right)\right|^2 + O(n\theta_n)$$

with

$$|\tilde{f}_{j,n}| \le \tilde{C}\theta_n,$$

$\tilde{C}$ being a constant independent both of n and j, and $n\theta_n = o(n)$. The latter displayed equality is a Riemann sum of a smooth function so that

$$\begin{aligned}\|\mathcal{P}_{U_n}(T_n(p))\|_F^2 &= \frac{n}{2\pi}\left(\frac{2\pi}{n}\sum_{j=0}^{n-1}\left| p\left(\frac{2\pi j}{n}\right)\right|^2\right) + o(n) \\ &= \frac{n}{2\pi}\left[\int_0^{2\pi} |p|^2 dt + O(n^{-1})\right] + o(n) \\ &= \frac{n}{2\pi}\int_0^{2\pi} |p|^2 dt + O(1) + o(n) \\ &= n\|p\|_{L^2}^2 + o(n)\end{aligned}$$

that is $\|E_2\|_F^2 = o(n)$, because of relation (19). Now we apply the result stated in Remark 5.4, with reference to the Frobenius norm, and again the singular values decomposition. Therefore, for every $\epsilon > 0$, we find sequences $\{\tilde{N}_{n,\epsilon}\}$ and $\{\tilde{R}_{n,\epsilon}\}$ such that

$$E_2 = \tilde{N}_{n,\epsilon} + \tilde{R}_{n,\epsilon}$$

$\|\tilde{N}_{n,\epsilon}\| < \epsilon/2$ and $\mathrm{rank}(\tilde{R}_{n,\epsilon}) \le \tilde{r}(\epsilon)n$ with

$$\lim_{\epsilon\to 0} \tilde{r}(\epsilon) = 0.$$

Finally we choose $\delta = \epsilon^2/4$ and the proof is concluded with reference to (18) and with the equalities

$$R_{n,\epsilon} = \hat{R}_{n,\epsilon^2/4} + \tilde{R}_{n,\epsilon}, \quad N_{n,\epsilon} = \hat{N}_{n,\epsilon^2/4} + \tilde{N}_{n,\epsilon}.$$

•

Remark 5.11. *We may notice that it is not necessary to know the function f to construct the continuous linear operator, but only its Fourier coefficients, which in turn can be numerically evaluated in a stable a fast way by using the (discrete) fast Fourier transform (see* [13]*).*

Remark 5.12. *It is interesting to notice that, when the symbol is not real-valued, the matrices $T_n(f)$ are rarely normal. However, a consequence of Theorem 5.10 (given in great generality in* [28]*) is that every sequence $\{T_n(f)\}$, f Lebesgue integrable, is close to a sequence of normal matrices (circulants) $\{\mathcal{P}_{U_n}(T_n(f))\}$, $U_n = F_n$ Fourier matrix, meaning that the difference sequence $\{T_n(f) - \mathcal{P}_{F_n}(T_n(f))\}$ is clustered to zero in the singular value sense.*

5.3.1 The check of the Korovkin test in the matrix Toeplitz setting

Let us follow Theorem 5.9 and let us consider first the most popular choice of the sequence $\{U_n\}$ that is the case where U_n is the discrete Fourier matrix F_n [37] with

$$F_n = \left(e^{-\sqrt{-1}\frac{2\pi jk}{n}}\right)_{j,k=0}^{n-1}.$$

In that case the related algebra $\mathcal{A}(F_n)$ (in the sense of Section 5.2) is the algebra of circulants [9, 13] in which $X \in \mathcal{A}(F_n)$ if and only if

$$X = \left(b_{(j-k) \bmod n}\right)_{j,k=0}^{n-1}$$

for some complex coefficients $b_0, \ldots, b_{n-1}$ where

$$X = F_n D(X) F_n^*, \quad [D(X)]_{j,j} = p_b\left(\frac{2\pi j}{n}\right), \quad p_b(t) = \sum_{k=0}^{n-1} b_k e^{\sqrt{-1}kt}.$$

In this case the minimization in the Frobenius norm is very easy because the problem decouples into n distinct one-variable minimization problems. In particular if $T = \left(a_{(j-k)}\right)_{j,k=0}^{n-1}$ is of Toeplitz type then

$$\min_{X\in\mathcal{A}(F_n)} \|T_X\|_F^2 = \min_{b_0,\ldots,b_{n-1}\in\mathbb{C}} \sum_{k=0}^{n-1} |a_k - b_k|^2(n-k) + |a_{k-n} - b_k|^2 k$$

whose optimal solution is expressible explicitly as

$$b_k = \frac{k a_{k-n} + (n-k) a_k}{n}. \qquad (20)$$

Now we are ready for the Korovkin test (see Theorem 5.9). We take $g(x) = e^{\sqrt{-1}x}$ and the Toeplitz matrix of size n generated by $g(x)$ that is $T_n(g(x))$ which is the transpose of the Jordan block of size n. In other words the (j,k) entry of $T_n(g(x))$ is 1 if $j-k=1$ and is zero otherwise. According to formula (20) we have $b_1 = 1-1/n$ and $b_k = 0$ for $k \neq 1$. Therefore the j-th eigenvalue of the optimal approximant is

$$\lambda_j(e^{\sqrt{-1}x}, n) = \left(F_n^* T_n(e^{\sqrt{-1}x}) F_n\right)_{j,j} = (1-1/n)e^{\sqrt{-1}\frac{2\pi j}{n}}.$$

As a consequence

$$\max_{j=0,\ldots,n-1} \|\lambda_j(e^{\sqrt{-1}x}, n) - e^{\sqrt{-1}x_j^{(n)}}\|_\infty = n^{-1}$$

with $x_j^{(n)} = \frac{2\pi j}{n}$ and the crucial assumption of Theorem 5.9 is fulfilled. Therefore, in the light of Theorem 5.9, we conclude that the sequence $\{T_n(f) - \mathcal{P}_{U_n}(T_n(f))\}$ is clustered at zero, in the strong sense, for every $f \in C_{2\pi}$ and is weakly clustered at zero for every $f \in L^1(0,2\pi)$ by virtue of Theorem 5.10.

In the literature we find many sequences of matrix algebras associated to fast transforms. The Korovkin test was performed originally for the circulants and all ω-circulants with $|\omega| = 1$ in [26]. The test in the case of all known sine and cosine algebras was verified in [10], while the case all Hartley transform spaces was treated in detail in [2]. It should be observed that the extension to the case of multilevel block structures (see [27]) is not difficult, as mentioned and briefly sketched in the next subsection.

5.3.2 The multilevel case

By following the notations of Tyrtyshnikov, a multilevel Toeplitz matrix of level N and dimension $n_1 \times n_2 \times \cdots \times n_N$ is defined as the matrix generated by the Fourier coefficients of a multivariate Lebesgue integrable function $f = f(x_1, \ldots, x_N)$ according to the law given in equations (6.1) at page 23 of [34] (see also [33]). In the following, for the sake of readability, we shall often write n for the N-tuple $(n_1, \ldots, n_N)$, $\hat{n} = n_1 \cdots n_N$. Then, for $f \in L^1((0,2\pi)^N)$ being N-variate and taking values into the rectangular matrices $\mathcal{M}_{s,t}(\mathbb{C})$, we define the $s\hat{n} \times t\hat{n}$ *multilevel Toeplitz matrix* by

$$T_n(f) = \sum_{j_1=-n_1+1}^{n_1-1} \cdots \sum_{j_N=-n_N+1}^{n_N-1} J_{n_1}^{(j_1)} \otimes \cdots \otimes J_{n_N}^{(j_N)} \otimes a_{(j_1,\ldots,j_N)}(f), \tag{21}$$

where $\otimes$ denotes the tensor or Kronecker product of matrices and $J_m^{(\ell)}$, $(-m+1 \leq \ell \leq m-1$, is the $m \times m$ matrix whose (i,j)th entry is 1 if $i-j=\ell$ and 0 otherwise; thus $\{J_{-m+1}, \ldots, J_{m-1}\}$ is the natural basis for the space of $m \times m$ Toeplitz matrices. In the usual multilevel indexing language, we say that $[T_n(f)]_{r,j} = a_{j-r}$ where $(1,\ldots,1) \leq j, r \leq n = (n_1, \ldots, n_N)$, i.e., $1 \leq j_\ell \leq n_\ell$ for $1 \leq \ell \leq N$ and the j-th Fourier coefficient of f that is

$$a_j = \frac{1}{(2\pi)^N} \int_{[0,2\pi]^N} f(t) e^{-\sqrt{-1}(j_1 t_1 + \cdots + j_N t_N)}\, dt,$$

$J = (j_1, \ldots, j_N)$, is a rectangular matrix of size $s \times t$. The latter relation together with (21) defines a sequence $\{T_n(\cdot)\}$ of operators, $T_n : L^1((0,2\pi)^N) \to$

$\mathcal{M}_{s\hat{n},t\hat{n}}(\mathbb{C})$, which are clearly linear, due to the linearity of the Fourier coefficients, and positive in a sense specified in [27]. Similarly, given the unitary matrix U_n related to the transform of a one-level algebra, a corresponding N-level algebra is defined as the set of $n_1 \times n_2 \times \cdots \times n_N$ matrices simultaneously diagonalized by means of the following tensor product of matrices

$$U_n = U_{n_1} \otimes U_{n_2} \otimes \cdots \otimes U_{n_N}, \tag{22}$$

with $A \otimes B$ being the matrix formed in block form as $(A_{i,j}B)$. Now, since we are interested in extending the results proved in the preceding sections to m dimensions, we analyze what is necessary to have and, especially, what is kept when we switch from one dimension to m dimensions: for the first level we used the Weierstrass and Korovkin Theorems, the last part of Lemma 5.5, Lemma 5.3 and Remark 5.4. Surprisingly enough, we find that all these tools hold or have a version in N dimensions: for the Korovkin and the Weierstrass Theorems, the multidimensional extensions are classic results, while Lemma 5.3 and Remark 5.4 contain statements not depending on the structure of the matrices and Lemma 5.5 is valid for any algebra and so for multilevel algebras as well (recall that U_n in (22) is unitary).

Therefore, we instantly deduce the validity in m dimensions of the main statements of this paper that is Theorems 5.8, 5.9, 5.9. However, we remark that we have no examples in which the strong convergence holds.

For instance, in the two-level circulant and τ cases, only the weak convergence has been proved because the number of the outliers is, in both cases, equal to $O(n_1+n_2)$ [6, 11] even if the function f is a bivariate polynomial: more precisely, this means that the hypotheses of Theorem 5.9, regarding the strong approximation in the polynomial case, are not fulfilled by the two-level circulant and τ algebras and therefore strong convergence cannot be proved in the general case (see also [27]). By using different tools, in [31] and [32] it has been proved that any sequence of preconditioners belonging to "Partially Equimodular" algebras [32] cannot be superlinear for sequence of multilevel Toeplitz matrices generated by simple positive polynomials. Here, "Partially Equimodular" refers to some very weak assumptions on U_n that are instanly fulfilled by all the known multilevel trigonometric algebras. In conclusion, the type of approximation delivered by the Frobenius optimal approximation and analyzed via the Korovkin theory, does not deliver the strong clustering in the multilevel setting: however, this drawback is intrinsic and hence such type of approximation is the best we can obtain as clearly shown in [31, 32] (see also [21, 29]).

6 Concluding Remarks

In these notes, we have shown classical applications of the Korovkin theory to the approximation of functions. A new view on the potential of the Korovkin Theorems is provided in a context of Numerical Linear Algebra. Here the main novelty is that the poor approximation from a quantitative viewpoint (see Exercise 3) is not a limitation. In reality, when employing preconditioned Krylov methods for the solution of large linear system, only a modest quantitative approximation of the spectrum is required in order to design optimal iterative solvers (see [12] and [4] for the specific commection between the spectrum of the preconditioned matrices and the iteration count of the associated iterative method).

7 Appendix A: General Tools

Linear and positives operators properties are necessary to prove the Korovkin Theorem. Therefore we consider following Lemma. It shows that the monotony is obtained by properties linearity and positivity.

Lemma 7.1. *Let $\mathcal{A}$ and $\mathcal{B}$ be vector spaces both endowed with a partial ordering and let $\Phi : \mathcal{A} \to \mathcal{B}$ be a linear and positive operator. Then Φ is monotone that is, if f and g are elements of $\mathcal{A}$ with $f \geq g$ we necessarily have $\Phi(f) \geq \Phi(g)$.*

Proof By the assumption $f \geq g$ we find

$$f - g \geq 0_{\mathcal{A}}.$$

then, using the positivity of the operator Φ we have

$$\Phi(f - g) \geq \Phi(0_{\mathcal{A}}) = 0_{\mathcal{B}}.$$

Finally by linearity we obtain $\Phi(f) - \Phi(g) \geq 0_{\mathcal{B}}$ which in turn leads to

$$\Phi(f) \geq \Phi(g).$$

□

As a direct consequence of the above lemma we get the isotonic property, at least if $f = f^*$. In fact if our spaces are equipped with the modulus we have

$$\pm f \leq |f|$$

and therefore the monotonicity implies

$$\pm\Phi(f) \leq \Phi(|f|) \Leftrightarrow |\Phi(f)| \leq \Phi(|f|).$$

The latter is known as isotonic property. However not all vectorial spaces have a modulus, for example a modulus is not defined for linear space generated by the test functions in Theorem 3.1.

Example: A well known example of isotonic functional is the integral defined in the space of complex valued integrable functions ($L^1[a, b]$). In fact for $f \in L^1[a, b]$ we know that

$$\left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx.$$

Concerning generalized versions of the isotonic property the reader is referred to [30], where, in the context of general matrix valued linear positive operators, it is proved that $|||\Phi_n(f)||| \leq |||\Phi_n(|f|)|||$ for every $f \in L^p(\Omega)$, $p \geq 1$, Ω equipped with a σ-finite measure, for every matrix valued linear positive operator $\Phi_n(\cdot)$, $n \geq 1$, from $L^p(\Omega)$ to $\mathcal{M}_n(\mathbb{C})$, for every unitarily invariant norm. The space L^p can be replaced by $C(K)$ for some compact set K in $\mathbb{R}^N$, $N \geq 1$, thanks to the Radon-Nykodim theorem (see [22]).

8 Appendix B: the (algebraic) Korovkin Theorem in N dimensions

We now give a full proof of the Korovkin Theorem in N dimensions. Concerning notation we recall that the symbol $\|\cdot\|_2$ indicates the Euclidean norm over $\mathbb{C}^N$ that is $\|x\|_2 = \left(\sum_{i=1}^N |x_i|^2\right)^{1/2}$ for $x \in \mathbb{C}^N$.

Theorem 8.1 (Korovkin)**.** *Let K be a compact set of $\mathbb{R}^N$ and let us consider the standard Banach space $C(K)$ endowed the sup-norm. Let us denote by $T = \{1, x_i, \|x\|_2^2 :\ i = 1, \ldots, N\}$ the standard Korovkin set of test functions and let us take a sequence $\{\Phi_n\}$ of linear positive operators from $C(K)$ in itself. If for any $g \in T$*

$$\Phi_n(g) \ \text{ uniformly converges to } g$$

then $\{\Phi_n\}$ is an approximation process i.e.

$$\Phi_n(f) \ \text{ uniformly converges to } f, \ \forall f \in C(K).$$

Proof We set $f \in C(K)$, we take an arbitrary $\epsilon > 0$ and we show that there exists $\bar{n}$ large enough for which $\|f - \Phi_n(f)\|_{\infty,K} \le \epsilon$, for every $n \ge \bar{n}$. Hence we take any point x of K and we consider the difference

$$f(x) - \Phi_n(f(y))(x)$$

where y is the dummy variable inside the operator Φ_n, where the function f acts. We now observe that the constant function 1 belongs to the "Korovkin test set " T and therefore $1 = (\Phi_n(1))(x) - \epsilon_n(1)(x)$, where $\epsilon_n(1)$ uniformly converges to zero over K. As a consequence, the use of the linearity of any Φ_n leads to

$$\begin{aligned} f(x) - \Phi_n(f(y))(x) &= [\Phi_n(1)(x) - \epsilon_n(1)(x)]f(x) - \Phi_n(f(y))(x) \\ &= \Phi_n(f(x) - f(y))(x) - \epsilon_n(1)(x)f(x). \end{aligned}$$

Thus there exists a value n_1 such that for every $n \ge n_1$ we find

$$|\epsilon_n(1)(x)f(x)| \le \|\epsilon_n(1)\|_{\infty,K}\|f\|_{\infty,K} \le \epsilon/4.$$

By exploiting the linearity and positivity of Φ_n (see Appendix A), we obtain

$$\begin{aligned} |f(x) - \Phi_n(f(y))(x)| &\le |\Phi_n(f(x) - f(y))(x)| + \epsilon/4 \\ &\le \Phi_n(|f(x) - f(y)|)(x) + \epsilon/4. \end{aligned} \tag{23}$$

As in the one-dimensional case the remaining part of the proof consists in a clever manipulation of the term $|f(x) - f(y)|$, of which we look for a "sharp" upper-bound, with the underlying idea of exploiting both the linearity and positivity of the operators and the assumption of convergence on the Korovkin test functions.

For manipulating the quantity $|f(x) - f(y)|$ we use the definition of uniform continuity. We are allowed to assume f uniformly continuous, since the notions of continuity and uniform continuity are the same when the function is defined on a compact set. Therefore for every $\epsilon_1 > 0$, there exists $\delta > 0$ such that $\|x - y\|_2 \le \delta$ implies $|f(x) - f(y)| \le \epsilon_1$. Now, depending on the new parameter

δ (and then in correspondence to the behavior of our function f), we define the pair of sets

$$Q_\delta = \{y \in K : \|x-y\|_2 \le \delta\}, \quad Q_\delta^C = K \backslash Q_\delta.$$

We observe that $|f(x)-f(y)|$ is bounded from above by ϵ_1 on Q_δ and by $2\|f\|_{\infty,K}$ su Q_δ^C, thanks to the triangle inequality. We denote by $\chi_{\mathcal{J}}$ the characteristic function of the set $\mathcal{J}$. Therefore the observation that $x, y \in Q_\delta^C$ implies $\|x-y\|_2 > \delta$ leads to

$$1 \le \|x-y\|_2^2/\delta^2,$$

for $x, y \in Q_\delta^C$. As a consequence we deduce the following chain of relationships.

$$\begin{aligned} |f(x)-f(y)| &\le \epsilon_1 \chi_{Q_\delta}(y) + 2\|f\|_{\infty,K}\chi_{Q_\delta^C}(y) \\ &\le \epsilon_1 \chi_{Q_\delta}(y) + 2\|f\|_{\infty,K}\chi_{Q_\delta^C}(y)\|x-y\|_2^2/\delta^2 \\ &\le \epsilon_1 + 2\|f\|_{\infty,K}\|x-y\|_2^2/\delta^2. \end{aligned}$$

The use of the positivity of Φ_n allows one to conclude

$$\Phi_n(|f(x)-f(y)|) \le \Phi_n\left(\epsilon_1 + 2\|f\|_{\infty,K}\|x-y\|_2^2/\delta^2\right).$$

Finally, using the linearity of Φ_n and setting $\Delta_n(f)(x) = |f(x) - \Phi_n(f(y))(x)|$, from (23) we infer the concluding relations, that is

$$\begin{aligned} \Delta_n(f)(x) &\le \Phi_n(|f(x)-f(y)|)(x) + \epsilon/4 \\ &\le \Phi_n\left(\epsilon_1 + 2\|f\|_{\infty,K}\|x-y\|_2^2/\delta^2\right)(x) + \epsilon/4 \\ &= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,K}/\delta^2\Phi_n\left(\|x-y\|_2^2\right)(x) + \epsilon/4 \\ &= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,K}/\delta^2 \\ &\quad \Phi_n\left(\sum_{i=1}^N x_i^2 - 2x_iy_i + y_i^2\right)(x) + \epsilon/4 \\ &= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,K}/\delta^2 \\ &\quad \sum_{i=1}^N \Phi_n\left(x_i^2 - 2x_iy_i + y_i^2\right)(x) + \epsilon/4 \\ &= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,K}/\delta^2 \\ &\quad \sum_{i=1}^N \left[x_i^2\Phi_n(1)(x) - 2x_i\Phi_n(y_i)(x) + \Phi_n(y_i^2)(x)\right] + \epsilon/4 \\ &= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,K}/\delta^2 \\ &\quad \left[\sum_{i=1}^N \left[x_i^2\Phi_n(1)(x) - 2x_i\Phi_n(y_i)(x)\right] + \Phi_n(\|y\|_2^2)(x)\right] \\ &\quad +\epsilon/4. \end{aligned}$$

We have now reduced the argument of Φ_n to a linear combination of test functions. Therefore we can proceed with explicit computations: setting $\Phi_n(g) =$

$g + \epsilon_n(g)$, with $\epsilon_n(g)$ uniformly converging to zero over K, we have

$$\begin{aligned}\Delta_n(f)(x) &\leq \epsilon_1(1+\epsilon_n(1)(x)) + 2\|f\|_{\infty,K}/\delta^2 \\ &\quad \left[\sum_{i=1}^N \left[x_i^2\Phi_n(1)(x) - 2x_i\Phi_n(y_i)(x)\right] + \Phi_n(\|y\|_2^2)(x)\right] + \epsilon/4 \\ &= \epsilon_1(1+\epsilon_n(1)(x)) + 2\|f\|_{\infty,K}/\delta^2 \\ &\quad \left[\sum_{i=1}^N \left[x_i^2\epsilon_n(1)(x) - 2x_i\epsilon_n(y_i)(x)\right] + \epsilon_n(\|y\|_2^2)(x)\right] + \epsilon/4.\end{aligned}$$

Finally, due to the uniform convergence to zero of the error functions $\epsilon_n(g)$, we deduce that there exists a value $\bar{n} \geq n_1$ for which, for every $n \geq \bar{n}$ we find $\epsilon_n(1))(x) \leq 1$ and

$$2\|f\|_{\infty,K}/\delta^2 \left[\sum_{i=1}^N \left[x_i^2\epsilon_n(1)(x) - 2x_i\epsilon_n(y_i)(x)\right] + \epsilon_n(\|y\|_2^2)(x)\right] \leq \epsilon/4.$$

In conclusion, by combining the partial results and by choosing properly ϵ_1 as a function of ϵ, we have proven the thesis namely

$$\Delta_n(f)(x) = |f(x) - \Phi_n(f(y))(x)| \leq \epsilon, \quad \forall n \geq \bar{n}.$$

•

9 Appendix C: the (periodic) Korovkin Theorem in N dimensions

We now give a full proof of the Korovkin Theorem in N dimensions in the 2π periodic case. We consider the Banach space

$$C_{2\pi} = \{f : \mathbb{R}^N \to \mathbb{C}, f \text{ continuous and periodic i.e. } f(x) = f(x \bmod 2\pi)\}$$

endowed with the sup-norm. The only substantial change will be in the set of test functions and in a clever variation of the notion of uniform continuity.

Theorem 9.1 (Korovkin). *Let us consider the standard Banach space $C_{2\pi}$ endowed the sup-norm. Let us denote by $T = \{1, e^{\sqrt{-1}x_i} : i = 1, \ldots, N\}$ the standard Korovkin set of 2π periodic test functions and let us take a sequence $\{\Phi_n\}$ of linear positive operators from $C_{2\pi}$ in itself. If for any $g \in T$*

$$\Phi_n(g) \text{ uniformly converges to } g$$

then $\{\Phi_n\}$ is an approximation process i.e.

$$\Phi_n(f) \text{ uniformly converges to } f, \ \forall f \in C_{2\pi}.$$

Proof. We fix $f \in C_{2\pi}$, we take an arbitrary $\epsilon > 0$ and we show that there exists $\bar{n}$ large enough for which $\|f - \Phi_n(f)\|_{\infty,K} \leq \epsilon$, for every $n \geq \bar{n}$. Hence we take any point x of K and we consider the difference

$$f(x) - \Phi_n(f(y))(x)$$

where y is the dummy variable inside the operator Φ_n, where the function f acts. We now observe that the constant function 1 belongs to the "Korovkin test set " T and therefore $1 = \Phi_n(1)(x) - \epsilon_n(1)(x)$, where $\epsilon_n(1)$ uniformly converges to zero over K. As a consequence, the use of the linearity of any Φ_n leads to

$$\begin{aligned} f(x) - \Phi_n(f(y))(x) &= [\Phi_n(1)(x) - \epsilon_n(1)(x)]f(x) - \Phi_n(f(y))(x) \\ &= \Phi_n(f(x) - f(y))(x) - \epsilon_n(1)(x)f(x). \end{aligned}$$

Thus there exists a value n_1 such that for every $n \geq n_1$ we find

$$|\epsilon_n(1)(x)f(x)| \leq \|\epsilon_n(1)\|_{\infty,K}\|f\|_{\infty,K} \leq \epsilon/4.$$

By exploiting the linearity and positivity of Φ_n (see Appendix A), we obtain

$$\begin{aligned} |f(x) - \Phi_n(f(y))(x)| &\leq |\Phi_n(f(x) - f(y))(x)| + \epsilon/4 \\ &\leq \Phi_n(|f(x) - f(y)|)(x) + \epsilon/4. \end{aligned} \tag{24}$$

As in Theorem 8.1, the remaining part of the proof consists in a clever manipulation of the term $|f(x) - f(y)|$, of which we look for a "sharp" upper-bound, with the underlying idea of exploiting both the linearity and positivity of the operators and the assumption of convergence on the Korovkin test functions.

For manipulating the quantity $|f(x) - f(y)|$ we use the definition of uniform continuity. We are allowed to assume f uniformly continuous, since the notions of continuity and uniform continuity are the same when the function is defined on a compact set and the set $\mathbb{R}^N$ can be reduced to the compact set $[0, 2\pi]^N$ thanks to the 2π-periodicity considered in the assumptions. Therefore for every $\epsilon_1 > 0$, there exists $\delta > 0$ such that $\|z(x) - z(y)\|_2 \leq \delta$ implies $|f(x) - f(y)| \leq \epsilon_1$ with

$$[z(x)]_i = e^{\sqrt{-1}x_i}, \quad i = 1, \ldots, N.$$

Now, depending on the new parameter δ (and then in correspondence to the behavior of our function f), we define the pair of sets

$$Q_\delta = \{y \in \mathbb{R}^N : \|z(x) - z(y)\|_2 \leq \delta\}, \quad Q_\delta^C = \mathbb{R}^N \backslash Q_\delta.$$

We observe that $|f(x) - f(y)|$ is bounded from above by ϵ_1 on Q_δ and by $2\|f\|_{\infty,\mathbb{R}^N}$ su Q_δ^C, thanks to the triangle inequality. We denote by $\chi_{\mathcal{J}}$ the characteristic function of the set $\mathcal{J}$. Therefore the observation that $x, y \in Q_\delta^C$ implies $\|z(x) - z(y)\|_2 > \delta$ leads to

$$1 \leq \|z(x) - z(y)\|_2^2/\delta^2,$$

for $x, y \in Q_\delta^C$. As a consequence we deduce the following chain of relationships.

$$\begin{aligned} |f(x) - f(y)| &\leq \epsilon_1\chi_{Q_\delta}(y) + 2\|f\|_{\infty,\mathbb{R}^N}\chi_{Q_\delta^C}(y) \\ &\leq \epsilon_1\chi_{Q_\delta}(y) + 2\|f\|_{\infty,\mathbb{R}^N}\chi_{Q_\delta^C}(y)\|z(x) - z(y)\|_2^2/\delta^2 \\ &\leq \epsilon_1 + 2\|f\|_{\infty,\mathbb{R}^N}\|z(x) - z(y)\|_2^2/\delta^2. \end{aligned}$$

The use of the positivity of Φ_n allows one to conclude

$$\Phi_n(|f(x) - f(y)|) \leq \Phi_n\left(\epsilon_1 + 2\|f\|_{\infty,\mathbb{R}^N}\|z(x) - z(y)\|_2^2/\delta^2\right).$$

Finally, using the linearity of Φ_n and setting $\Delta_n(f)(x) = |f(x) - \Phi_n(f(y))(x)|$, from (24) we infer the concluding relations, that is

$$\begin{aligned}
\Delta_n(f)(x) &\leq \Phi_n(|f(x) - f(y)|)(x) + \epsilon/4 \\
&\leq \Phi_n\left(\epsilon_1 + 2\|f\|_{\infty,\mathbb{R}^N}\|z(x) - z(y)\|_2^2/\delta^2\right)(x) + \epsilon/4 \\
&= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,\mathbb{R}^N}/\delta^2\Phi_n\left(\|z(x) - z(y)\|_2^2\right)(x) + \epsilon/4 \\
&= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,\mathbb{R}^N}/\delta^2 \\
&\quad \Phi_n\left(\sum_{i=1}^N 1 - e^{-\sqrt{-1}x_i}e^{\sqrt{-1}y_i} - e^{\sqrt{-1}x_i}e^{-\sqrt{-1}y_i} + 1\right)(x) + \epsilon/4 \\
&= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,\mathbb{R}^N}/\delta^2 \\
&\quad \sum_{i=1}^N \Phi_n\left(2 - e^{-\sqrt{-1}x_i}e^{\sqrt{-1}y_i} - e^{\sqrt{-1}x_i}e^{-\sqrt{-1}y_i}\right)(x) + \epsilon/4 \\
&= \epsilon_1\Phi_n(1)(x) + 2\|f\|_{\infty,\mathbb{R}^N}/\delta^2 \\
&\quad \sum_{i=1}^N \left[\Phi_n(2)(x) - [z(x)]_i\Phi_n(\overline{[z(y)]_i})(x) - \overline{[z(x)]_i}\Phi_n([z(y)]_i)(x)\right] + \\
&\quad +\epsilon/4.
\end{aligned}$$

We have now reduced the argument of Φ_n to a linear combination of test functions. Therefore we can proceed with explicit computations: setting $\Phi_n(g) = g + \epsilon_n(g)$, with $\epsilon_n(g)$ uniformly converging to zero over K, we obtain that $\Delta_n(f)(x)$ is bounded from above by

$$\begin{aligned}
&\epsilon_1(1 + \epsilon_n(1)(x)) + 2\|f\|_{\infty,\mathbb{R}^N}/\delta^2 \cdot \\
&\cdot\left[\sum_{i=1}^N \left[2\epsilon_n(1)(x) - [z(x)]_i\overline{\epsilon_n([z(y)]_i)}(x) - \overline{[z(x)]_i}\epsilon_n([z(y)]_i)(x)\right]\right] + \\
&+\epsilon/4.
\end{aligned}$$

Finally, due to the uniform convergence to zero of the error functions $\epsilon_n(g)$, we deduce that there exists a value $\bar{n} \geq n_1$ for which, for every $n \geq \bar{n}$ we find $\epsilon_n(1)(x)) \leq 1$ and the quantity

$$2\|f\|_{\infty,\mathbb{R}^N}/\delta^2\left[\sum_{i=1}^N 2\epsilon_n(1)(x) - [z(x)]_i\overline{\epsilon_n([z(y)]_i)}(x) - \overline{[z(x)]_i}\epsilon_n([z(y)]_i)(x)\right]$$

globally bounded from above by $\epsilon/4$. In conclusion, by combining the partial results and by choosing properly ϵ_1 as a function of ϵ, we have proven the thesis namely

$$\Delta_n(f)(x) = |f(x) - \Phi_n(f(y))(x)| \leq \epsilon, \quad \forall n \geq \bar{n}.$$

•

10 Exercises

1. With reference to the notations of Section 4.3 complete the proof of the Weierstrass Theorem by using the Gauss-Weierstrass operators.

2. Let $\{\Phi_n\}$ be a sequence of linear positive operators from $C[0,1]$ into itself such that $\Phi_n(g(t))(x) = g(x)$ for all $x \in [0,1]$, for all standard test functions $g(t) = t^j$, $j = 0,1,2$. Prove that the sequence is trivial i.e. $\Phi_n(f) \equiv f$ for every $f \in C[0,1]$. (Compare this result with the case of interpolation operators: conclude that interpolation operators, which are linear, cannot be positive).

3. Let $\{\Phi_n\}$ be a sequence of linear positive operators such that Φ_n acts on $C[0,1]$, $\Phi_n(f)$ is a polynomial of degree at most n for every $f \in C[0,1]$ and such that the infinity norm of $n^2(\Phi_n(g(t))(x) - g(x))$ goes to zero as n tends to infinity, for all standard test functions $g(t) = t^j$, $j = 0,1,2$. Prove that such a sequence cannot exist. Hint: for that exercise a good starting point is the following saturation result. Let $f \in C^k[0,1]$ for some $k \geq 0$ and let
$$E_n(f) = \min_{p \in \mathbb{P}_n} \|f - p\|_{\infty,[0,1]}.$$
Then for every $k \geq 0$, for every non-increasing, nonnegative sequence ϵ_n tending to zero as n tends to infinity, there exists $f \in C^k[0,1]$ such that
$$E_n(f) n^k \geq \epsilon_n.$$
(Compare this result with the case of interpolation operators and especially with the Faber Theorem [35] and Jackson estimates [16, 19]: interpolation operators do not guarantee convergence for every continuous function f, but could be extremely fast convergent if f is smoother).

4. Give a (constructive) proof of the saturation theorems stated in the previous exercise: start with the case $k = 0$.

5. With reference to the notations of Section 4.3, consider the sentence "the proof is only partially constructive since $p_{n,m}$ is defined using information on f not available in general (except via numerical evaluations)" given in the last part of the section: identify this information.

6. Prove the multiplicative separability property for the Bernstein operators i.e.
$$B_n(f(t))(x) = B_{\tilde{n}}(g(\tilde{t}))(\tilde{x}) B_{\hat{n}}(h(\hat{t}))(\hat{x})$$
whenever $f(t) = g(\tilde{t})h(\hat{t})$ with
$$\begin{aligned}
n &= (n_1, \ldots, n_N), \\
t &= (t_1, \ldots, t_N), \\
x &= (x_1, \ldots, x_N), \\
\tilde{v} &= (v_{i_1}, \ldots, i_q), \quad 1 \leq i_1 < \cdots < i_q \leq N, \\
\hat{v} &= (v_{j_1}, \ldots, j_r), \quad 1 \leq j_1 < \cdots < j_r \leq N, \\
q + r &= N, \\
& \quad \{i_1, \ldots, i_q\} \cup \{j_1, \ldots, j_r\} = \{1, \ldots, N\}.
\end{aligned}$$

7. Verify the multidimensional Korovkin test for multidimensional Bernstein operators (use the multiplicative separability property of the previous exercise and reduce the multidimensional case to the unidimensional case already treated in Section 4.1).

8. Prove the Korovkin Theorem by varying some ingredients

   a "Following" the proof of Theorem 9.1, prove:

   $$\Phi_n(g) \text{ converges in } L^1 \text{ norm to } g, \ \forall g \in T$$

   then the sequence $\{\Phi_n\}$ of LPOs is an approximation process namely

   $$\Phi_n(f) \text{ converges in } L^1 \text{ norm to } f, \ \forall f \in C_{2\pi}.$$

   b Give the "Korovkin test" set in the case periodic and real.

   c Give the "Korovkin test" set in the case periodic, real, and even i.e. $f(-x) = f(x)$, $x \in [0, \pi]$.

   d With reference to the previous items give the "Korovkin test" set in the multidimensional setting.

9. For $f \in L^1(-\pi, \pi)$ let $C_n(f)(x) = \frac{1}{n+1} \sum_{k=0}^{n} F_k(f)(x)$ be its Cesaro sum of degree n with $F_k(f)(x) = \sum_{j=-k}^{k} a_j e^{\sqrt{-1}jx}$, being its k-th Fourier sum and $a_j = a_j(f)$ being its j-th Fourier coefficient. Let $T_n(f)$ be the $(n+1) \times (n+1)$ Toeplitz matrix defined as

   $$T_n(f) = (a_{s-t})_{s,t=0}^{n}.$$

   and $v(x)$ be the vector of length $n+1$ with k-th entry given by $e^{-\sqrt{-1}kx}$.
   [a] Prove the identity $C_n(f)(x) = \frac{1}{n+1} v^*(x) T_n(f) v(x)$.
   [b] We already proved that $T_n(\cdot) : L^1(-\pi, \pi) \to M_n(\mathbf{C})$ is a LPO. Using the latter prove that $C_n(\cdot) : L^1(-\pi, \pi) \to L^1(-\pi, \pi)$ is a LPO.
   [c] Assuming that $f \in L^\infty(-\pi, \pi)$ and that f is real-valued, prove that $C_n(f)(x) \geq \operatorname{ess\,inf} f$ and $C_n(f)(x) \leq \operatorname{ess\,sup} f$, $\forall x \in [-\pi, \pi]$ (if you prefer, assume that $f \in C_{2\pi}$).
   [d] By using the right Korovkin Theorem prove that $\forall f \in \mathcal{C}_{2\pi}$

   $$\lim_{n \to \infty} \|C_n(f)(x) - f(x)\|_\infty = 0.$$

   Notice that the latter furnishes a constructive proof of the second Weierstrass Theorem (see Theorem 1.2).
   [e] From the monumental book of Zygmund on Trigonometric Series [41] we learn that the k-th Cesaro sum is such that

   $$C_k(f)(x) = f(x) + \frac{s(f)(x)}{k+1} + r_k(f)(x) \qquad (25)$$

   where f is smooth enough, $s(f)(x)$ is a bounded function independent of k, and $\|r_k(f)(x)\|_\infty$ is of the order of the best trigonometric approximation of degree k, for $k \to \infty$. Suppose that one has computed $g_1 = C_{n-1}(f)(x)$ and $g_2 = C_{3n-1}(f)(x)$ for some sufficiently large value of n. How to exploit (25) in order to compute $f(x)$ within an error of the order of $\|r_{n-1}(f)(x)\|_\infty$?
   [f] Show that $C_n(2 - 2\cos(t))(0) \sim n^{-1}$ by direct computation.
   We set $h(t) = (e^{t^2} - 1)(3 + \operatorname{sign}(|t| - 1)|\sin(t+1)|^{5/\pi})$, $t \in (-\pi, \pi]$.
   Is it true that $C_n(h(t))(0) \sim n^{-1}$?
   Is it true that $C_n(h^\alpha(t))(0) \sim n^{-1}$ for $\alpha \in (1, 2]$?
   Is it true that $C_n(h^\alpha(t))(0) \sim n^{-1}$ for every fixed α?

10. Prove all Korovkin results discussed so far, namely Theorems 8.1, 9.1, and Lemma 5.6, by using the size depending norms

$$\|f\|_{\infty,J_n} = \sup_{x\in J_n} |f(x),$$

where $J_n \subset K$, K compact, in the algebraic case and $J_n \bmod 2\pi \cdot e \subset [0, 2\pi)^N$, $N \geq 1$, in the periodic case with e being the N dimensional vector of all ones and with $x \in J_n \bmod 2\pi \cdot e$ if x_j, $j = 1, \ldots, N$, is of the form $y_j \bmod 2\pi$ and $y \in J_n$. The above generalization was used in [26, 27]; refer to [1] for a monumental survey on the Korovkin theory.

11. With reference to Lemma 5.6 prove that the same statement holds true if $\mathcal{G}$ is the linear space of the continuous even periodic functions, if p is an even polynomial (with a finite cosine expansion) and if the Chebyshev set $\{p_i\}_{i=0,1,2}$ is given by $1, \cos x, \cos 2x$ over $I = [0, \pi]$ with $J_n \subset I$ (use a similar argument and combine it with a simple technique of inductive type).

12. Let Ω a set of $\mathbb{R}^d$ with $d \geq 2$. Let $C = \{g_0(x), \ldots, g_k(x)\}$, $k \geq 1$, be a collection of real-valued continuous functions over Ω. Prove that there exist a collection of pair-wise distinct points $\{x_0, \ldots, x_k\}$ (i.e. $x_i \neq x_j$ if $i \neq j$) such that the $(k+1) \times (k+1)$ Vandermonde-like matrix

$$(g_j(x_k))_{j,k=0}^{k}$$

is singular whenever Ω contains a "Y shaped" domain (the tram switching proof).

References

[1] F. Altomare, M. Campiti, (Appendix A by Michael Pannenberg and Appendix B by Ferdinand Beckhoff), *Korovkin-type approximation theory and its applications.* de Gruyter Studies in Mathematics, 17. Walter de Gruyter, Berlin, 1994.

[2] A. Arico', S. Serra-Capizzano, M. Tasche, "Fast and numerically stable algorithms for discrete Hartley transforms and applications to preconditioning", *Comm. Informat. Syst.*, **5-1** (2005), pp. 21–68.

[3] O. Axelsson, V. Barker, *Finite Element Solution of Boundary Value Problems, Theory and Computation.* Academic Press Inc., New York, 1984.

[4] O. Axelsson, G. Lindskög, "The rate of convergence of the preconditioned conjugate gradient method", *Numer. Math.*, **52** (1986), pp. 499–523.

[5] R. Bhatia. *Matrix Analysis.* Springer-Verlag, New York, NY, 1997.

[6] R.H. Chan, X. Jin, "A family of block preconditioners for block systems", *SIAM J. Sci. Stat. Comput.*, **13** (1992), pp. 1218–1235.

[7] R.H. Chan, M. Ng, "Conjugate gradient method for Toeplitz systems", *SIAM Rev.*, **38** (1996), 427–482.

[8] T.F. Chan, "An optimal circulant preconditioner for Toeplitz systems", *SIAM J. Sci. Stat. Comput.*, **9** (1988), pp. 766–771.

[9] P. Davis, *Circulant Matrices.* John Wiley and Sons, New York, 1979.

[10] F. Di Benedetto, S. Serra-Capizzano, "A unifying approach to abstract matrix algebra preconditioning", *Numer. Math.*, **82-1** (1999), pp. 57–90.

[11] F. Di Benedetto, S. Serra-Capizzano, "Optimal multilevel matrix algebra operators", *Linear Multilin. Algebra*, **48** (2000), pp. 35–66.

[12] F. Didone, M. Giannuzzi, L. Paulon, S. Serra-Capizzano, *Solution of a large linear system using the Preconditioned Conjugate Gradient Method.* From lectures titled "Approssimazione matriciale, precondizionamento e sistemi lineari di grandi dimensioni" by S. Serra-Capizzano in Rome, at the Department "MeMoMat", University "La Sapienza", 2010.

[13] F. Didone, M. Giannuzzi, L. Paulon, S. Serra-Capizzano, *How to perform a Toeplitz matrix-vector product efficiently.* From lectures titled "Approssimazione matriciale, precondizionamento e sistemi lineari di grandi dimensioni" by S. Serra-Capizzano in Rome, at the Department "MeMoMat", University "La Sapienza", 2010.

[14] G.H. Golub, C.F. Van Loan, *Matrix Computations.* The Johns Hopkins University Press, Baltimore 1983.

[15] A. Greenbaum, *Iterative methods for solving linear systems*, SIAM publ., Philadelphia, 1997.

[16] D. Jackson, *The Theory of Approximation.* American Mathematical Society, New York, 1930.

[17] P.P. Korovkin, *Linear Operators and Approximation Theory* (English translation). Hindustan Publishing Co., Delhi, 1960.

[18] K. Kumar, N. Namboodiri, S. Serra-Capizzano, *A tutorial on the Korovkin theory in approximation, (numerical) linear algebra, and operator theory.* In preparation.

[19] I.P. Natanson, *Constructive Function Theory, I.* Frederick Ungar Publishing Co., New York, 1964.

[20] M. Ng, *Iterative methods for Toeplitz systems.* Oxford University Press, New York, 2004. .

[21] D. Noutsos, S. Serra Capizzano, P. Vassalos", "Matrix algebra preconditioners for multilevel Toeplitz systems do not insure optimal convergence rate", *Theoret. Comput. Sci.*, **315** (2004), pp. 557–579.

[22] W. Rudin, *Real and Complex Analysis.* McGraw-Hill, Singapore, 1986.

[23] W. Rudin, *Functional Analysis.* McGraw-Hill, Singapore, 1986.

[24] Y. Saad, *Iterative methods for sparse linear systems.* PWS Publishing, Boston, MA, 1996.

[25] S. Serra-Capizzano, "Some theorems on linear positive operators and functionals and their applications", *Comput. Math. Appl.*, **39-7/8** (2000), pp. 139–167.

[26] S. Serra-Capizzano, "A Korovkin-type Theory for finite Toeplitz operators via matrix algebras", *Numer. Math.*, **82-1** (1999), pp. 117–142.

[27] S. Serra-Capizzano, "A Korovkin based approximation of multilevel Toeplitz matrices (with rectangular unstructured blocks) via multilevel trigonometric matrix spaces", *SIAM J. Numer. Anal.*, **36-6** (1999), pp. 1831–1857.

[28] S. Serra-Capizzano, "Korovkin tests, approximation, and ergodic theory", *Math. Comput.*, **69** (2000), pp. 1533–1558.

[29] S. Serra-Capizzano, "Matrix algebra preconditioners for multilevel Toeplitz matrices are not superlinear", *Linear Algebra Appl.*, **343/344** (2002), pp. 303–319.

[30] S. Serra-Capizzano, P. Tilli, "On unitarily invariant norms of matrix valued linear positive operators", *J. Inequalities Appl.*, **7-3** (2002), pp. 309–330.

[31] S. Serra Capizzano, E. Tyrtyshnikov, "Any circulant-like preconditioner for multilevel matrices is not superlinear", *SIAM J. Matrix Anal. Appl.,* **21-2** (1999), pp. 431–439.

[32] S. Serra-Capizzano, E. Tyrtyshnikov, "How to prove that a preconditioner can not be superlinear", *Math. Comput.*, **72-243** (2003), pp. 1305–1316.

[33] P. Tilli, "A note on the spectral distribution of Toeplitz matrices", *Linear Multilin. Algebra*, **45** (1998), pp. 147–159.

[34] E. Tyrtyshnikov, "A unifying approach to some old and new theorems on distribution and clustering", *Linear Algebra Appl.*, **232** (1996), pp. 1–43.

[35] E. Tyrtyshnikov, *A brief introduction to numerical analysis.* Birkhäuser Boston, Inc., Boston, MA, 1997.

[36] E. Tyrtyshnikov, N. Zamarashkin, "Spectra of multilevel Toeplitz matrices: advanced theory via simple matrix relationships", *Linear Algebra Appl.*, **270** (1998), pp. 15–27.

[37] C. Van Loan, *Computational Frameworks for the Fast Fourier Transform.* SIAM, Philadelphia, 1992.

[38] H. Widom, *Toeplitz matrices.* In Studies in real and complex analysis, I. Hirshman Jr. Ed., Math. Ass. Amer., 1965.

[39] H. Widom, "Szegö limit Theorem: the higher dimensional matrix case", *J. Funct. Anal.*, **39** (1980), pp. 182–198.

[40] H. Widom, "On the singular values of Toeplitz matrices", *Zeit. Anal. Anw.*, **8** (1989), pp. 221–229.

[41] A. Zygmund, *Trigonometric Series.* Cambridge University Press, Cambridge, 1959.