A note on the (regularizing) preconditioning of $g$-Toeplitz sequences via $g$-circulants

Claudio Estatico$^{a}$, Eric Ngondiep$^{a}$, Stefano Serra-Capizzano$^{a,\ast}$, Debora Sesana$^{b,a}$

$^{a}$ Dipartimento di Fisica e Matematica, Università dell’Insubria, Via Valleggio 11, I-22100 Como, Italy
$^{b}$ Dipartimento di Scienze Economiche e Metodi Quantitativi, Università del Piemonte Orientale, Via Perrone 18, I-28100 Novara, Italy

**ARTICLE INFO**

Article history:
Received 30 December 2010

MSC:
65F10
15A18

Keywords:
g-circulants
g-Toeplitz
Spectral distributions
Clustering
Preconditioning
Multigrid methods

**ABSTRACT**

For a given nonnegative integer $g$, a matrix $A_n$ of size $n$ is called $g$-Toeplitz if its entries obey the rule $A_n = [a_{r-gs}]_{r,s=0}^{n-1}$. Analogously, a matrix $A_n$ again of size $n$ is called $g$-circulant if $A_n = [a_{(r-gs)\mod n}]_{r,s=0}^{n-1}$. In a recent work we studied the asymptotic properties, in terms of spectral distribution, of both $g$-circulant and $g$-Toeplitz sequences in the case where $[a_k]$ can be interpreted as the sequence of Fourier coefficients of an integrable function $f$ over the domain $(-\pi, \pi)$. Here we are interested in the preconditioning problem which is well understood and widely studied in the last three decades in the classical Toeplitz case, i.e., for $g = 1$. In particular, we consider the generalized case with $g \geq 2$ and the nontrivial result is that the preconditioned sequence $\{P_n\} = \{P_n^{-1}A_n\}$, where $P_n$ is the sequence of preconditioner, cannot be clustered at 1 so that the case of $g = 1$ is exceptional. However, while a standard preconditioning cannot be achieved, the result has a potential positive implication since there exist choices of $g$-circulant sequences which can be used as basic preconditioning sequences for the corresponding $g$-Toeplitz structures. Generalizations to the block and multilevel case are also considered, where $g$ is a vector with nonnegative integer entries. A few numerical experiments, related to a specific application in signal restoration, are presented and critically discussed.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

A matrix $A_n$ of size $n$ is called $g$-Toeplitz if its entries obey the rule $A_n = [a_{r-gs}]_{r,s=0}^{n-1}$, where $g$ is a nonnegative integer. A matrix $A_n$ of size $n$ is called $g$-circulant if $A_n = [a_{(r-gs)\mod n}]_{r,s=0}^{n-1}$: for an introduction and for the algebraic properties of such matrices we refer to Section 5.1 of the classical book by Davis [1], while new additional results can be found in [2] and references therein. On the other hand, such structured matrices are encountered in many fields such as e.g. multigrid methods [3,4], wavelet analysis [5], and subdivision algorithms or, equivalently, in the associated refinement equations, see [6] and references therein. Furthermore, it is instructive to recall that Gilbert Strang [7] has shown rich connections between dilation equations in the wavelet context and multigrid algorithms [3,8], when constructing the restriction/prolongation operators [9] with various boundary conditions. It is worth noticing that the use of different boundary conditions is quite natural when dealing with signal/image restoration problems or differential equations, see [10,11].

---

* Corresponding author.
E-mail addresses: claudio.estatico@uninsubria.it (C. Estatico), eric.ngondiep@uninsubria.it (E. Ngondiep), stefano.serrac@uninsubria.it (S. Serra-Capizzano), debora.sesana@eco.unipmn.it (D. Sesana).

0377-0427/$–$ see front matter © 2011 Elsevier B.V. All rights reserved.
In a recent paper [12] we addressed the problem of characterizing the singular values of $g$-circulants and of providing an asymptotic analysis of the distribution results for the singular values of $g$-Toeplitz sequences, in the case where the sequence of values $(a_n)$, defining the entries of the matrices, can be interpreted as the sequence of Fourier coefficients of an integrable function $f$ over the domain $(-\pi, \pi)$. Such results were plainly generalized to the block, multilevel case, amounting to choosing the symbol $f$ multivariate, i.e., defined on the set $(-\pi, \pi)^d$ for some $d > 1$, and matrix-valued, i.e., such that $f(x)$ is a matrix of given size $p \times q$.

Here we consider the preconditioning problem. In particular, we consider the general case with $g \geq 2$ and the interesting result is that the preconditioned sequence $\{\mathcal{F}_k\} = \{P_{n-k} A_n\}$, where $\{P_n\}$ is the sequence of preconditioner, cannot be clustered at 1 so that the case of $g = 1$, widely studied in the literature, is exceptional (see e.g. [13,14] for the one-level case, [15] for the multilevel case, and [16] for the multilevel block case). However, while the optimal preconditioning cannot be achieved, the result has a potential positive implication since there exist choices of $g$-circulant sequences which are regularizing preconditioning sequences for the corresponding $g$-Toeplitz structures. Indeed the nice feature is that, as required in a regularization context, the singular values of $\{\mathcal{F}_n\}$ related to the subspaces of non-negligible singular values of $A_n$ are well clustered at unity. On the other hand, in the subspaces where $A_n$ is highly contractive, the action of the preconditioning sequence is negligible (in these subspaces the original operator is ill-posed). These two facts are very welcome, but a negative aspect which can be understood also from the numerical experiments is that the subspaces associated with degenerating singular values has a non-trivial intersection with the low frequencies where usually the signals/images live and the latter is independent of the analytical features of the symbol $f$. Hence the problems $A_n x = b$ are inherently highly ill-posed and this cannot be changed by any choice of the preconditioning sequence. Anyway, $g$-circulant sequences of preconditioners can be easily regularized by means of fast numerical linear algebra based on trigonometric fast transformations like FFT, so that their regularized version can be suitably used for $g$-Toeplitz preconditioning. Generalizations to the block and multilevel case are also considered. Numerical results, also instructive for specific applications in image deblurring and denoising, are presented and critically discussed. In particular they confirm the regularizing features of the proposed preconditioners in their regularized versions, even in the presence of an extreme ill-posedness of the sequence of algebraic problems.

The paper is organized as follows. In Section 2 we introduce useful definitions and well-known results concerning the notion of spectral distribution, while Section 3 is devoted to some preparatory and general results on preconditioning and clustering. In Section 4 we report distribution results on $g$-circulants and $g$-Toeplitz sequences. Section 5 is devoted to the preconditioning analysis both in the standard and regularizing sense, while in Section 5.5 we discuss the generalization of the results when we deal with the multilevel block case. The aim of Section 6 is to present a few numerical experiments confirming the theoretical findings, while in Section 7 we draw conclusions and indicate future lines of research.

2. General definitions and tools from spectral distribution theory

For any function $F$ defined on $\mathbb{R}_+^d$ and for any $n \times m$ matrix $A$, the symbol $\Sigma_n (F, A)$ stands for the mean

$$\Sigma_n (F, A) := \frac{1}{\min[n,m]} \sum_{j=1}^{\min[n,m]} F(\sigma_j(A)),$$

where $\sigma_j(A)$, for $j = 1, \ldots, \min[n,m]$ are the singular values of $A$.

Throughout this paper we speak also of matrix sequences as sequences $\{A_k\}$ where $A_k$ is an $n(k) \times m(k)$ matrix with $\min[n(k), m(k)] \to \infty$ as $k \to \infty$. When $n(k) = m(k)$, that is all the involved matrices are square, and this will occur often in the paper, we will not need the extra parameter $k$ and we will consider simply matrix sequences of the form $\{A_n\}$.

Concerning the case of matrix sequences an important notion is that of the spectral distribution in the eigenvalue or singular value sense, linking the collective behavior of the eigenvalues or singular values of all the matrices in the sequence to a given function (or to a given measure). The notion goes back to Weyl and has been investigated by many authors in the Toeplitz and locally Toeplitz context (see the book by Böttcher and Silbermann [17] where many classical results by the authors, Szegő, Avram, Parter, Widom, Tytunshikov, and many others can be found, and more recent results in [18–25]). Here we treat the notion of spectral distribution only in the singular value sense since our analysis is devoted to singular values: regarding the eigenvalue distribution, both in the preconditioned and even non-preconditioned case, is substantially trickier given the inherent non-normality of the involved matrices.

**Definition 2.1.** Let $C_0(\mathbb{R}_+^d)$ be the set of continuous functions with bounded support defined over the nonnegative real numbers, $d$ a positive integer, and $\theta$ a complex-valued measurable function defined on a set $G \subset \mathbb{R}_+^d$ of finite and positive Lebesgue measure $m(G)$. Here $G$ will often be equal to $(-\pi, \pi)^d$ so that $e^{i\theta} = e^{i\theta} \mathbb{T}$ with $\mathbb{T}$ denoting the complex unit circle. A matrix sequence $\{A_k\}$ is said to be distributed (in the sense of the singular values) as the pair $(\theta, G)$, or to have the distribution function $\theta$ (denoted as $\{A_k\} \sim_{\sigma} (\theta, G)$), if $\forall F \in C_0(\mathbb{R}_+^d)$, the following limit relation holds

$$\lim_{k \to \infty} \Sigma_n (F, A_k) = \frac{1}{m(G)} \int_G F(\theta(t)) \, dt, \quad t = (t_1, \ldots, t_d).$$
When considering \( \theta \) taking values in \( M_{pq} \), where \( M_{pq} \) is the space of \( p \times q \) matrices with complex entries and a function is considered to be measurable if and only if the component functions are, we say that \( \{A_k\} \sim_\sigma(\theta, G) \) when for every \( F \in C_0(\mathbb{R}^n) \) we have

\[
\lim_{k \to \infty} \Sigma_\sigma(F, A_k) = \frac{1}{m[G]} \int_{G} \sum_{j=1}^{\min(p,q)} F(\sigma_j(\theta(t))) \frac{dt}{\min(p,q)}, \quad t = (t_1, \ldots, t_d),
\]

with \( \sigma_j(\theta(t)) = \sqrt{\lambda_j(\theta(t)\theta^*(t))} \). Finally we say that two sequences \( \{A_k\} \) and \( \{B_k\} \) are \emph{equally distributed} in the sense of singular values if, \( \forall F \in C_0(\mathbb{R}^q) \), we have

\[
\lim_{k \to \infty} [\Sigma_\sigma(F, B_k) - \Sigma_\sigma(F, A_k)] = 0.
\]

**Definition 2.2** ([26]). Consider a sequence of matrices \( \{A_n\} \), where \( A_n \) of size \( d_n \), and a set \( M \) in the nonnegative real line. Take \( \varepsilon > 0 \) and denote by \( M_\varepsilon \) the \( \varepsilon \)-extension of \( M \), i.e. the union of all real \( \varepsilon \)-balls encircling \( M \)'s points. For any \( n \), let \( \gamma_n(\varepsilon) \) be the number of those singular values of \( A_n \) not belonging to \( M_\varepsilon \). Then \( M \) is called a \emph{general singular value cluster} if \( \forall \varepsilon > 0 \)

\[
\lim_{n \to \infty} \frac{\gamma_n(\varepsilon)}{d_n} = 0
\]

and \( M \) is called a \emph{proper singular value cluster} if \( \forall \varepsilon > 0 \)

\[
\gamma_n(\varepsilon) \leq c(\varepsilon),
\]

where \( c(\varepsilon) \) is independent of \( n \). In the case where \( M = \{p\} \) then we simply say that \( \{A_n\} \) is \emph{clustered} at \( p \) with respect to the singular values.

**Proposition 2.1** ([27,28]). If \( \{A_n\} \), \( \{B_n\} \), and \( \{Q_n\} \) are sequences of matrices of strictly increasing dimensions \( \{d_n\} \), such that \( \{A_n\} \sim_\sigma(\theta, G) \), \( \{B_n\} \sim_\sigma(0, G) \), and \( \|Q_n\|_2 \leq M \) for some nonnegative constant \( M \) independent of \( n \), where \( \| \cdot \|_2 \) denotes the spectral 2-norm, then

\[
\{A_n + B_n\} \sim_\sigma(\theta, G),
\]

\[
\{B_nQ_n\} \sim_\sigma(0, G),
\]

\[
\{Q_nB_n\} \sim_\sigma(0, G).
\]

3. General definitions and tools from preconditioning theory

When preconditioning a spectrally bounded sequence it is compulsory to use a spectrally bounded sequence of preconditioners; otherwise the preconditioned sequence will have necessarily the minimal singular value tending to zero with the size and this is known to spoil the convergence speed of any Krylov-like technique (see for instance the classical result of Axelsson and Lindhög [29] in the context of the conjugate gradient). Therefore, if we look at a preconditioned sequence \( \{P_n\} = \{P_n^{-1}A_n\} \), where \( \{P_n\} \) is the sequence of preconditioners, such that \( \{P_n - I_n\} \) is clustered at 0, then the difference between the original sequence and the sequence of preconditioners, that is \( \{A_n - P_n\} \), should be clustered at zero too. The latter tells us that if the original sequence has a given distribution then, necessarily, the preconditioning sequence has to be chosen with the same distribution. Such key statements and other theoretical tools are given and proven in the next subsection.

3.1. Tools and machineries

In this section, first we give some basic definitions and we introduce some general tools for the spectral analysis of matrix sequences. As already mentioned in the previous section, by \( \{d_n\} \) we denote an increasing sequence of natural numbers.

**Definition 3.1.** A sequence of matrices \( \{X_n\} \), with \( X_n \) of size \( d_n \), is said to be \emph{sparsely vanishing} if there exists a nonnegative function \( x(s) \) with \( \lim_{s \to 0} x(s) = 0 \) so that \( \forall \varepsilon > 0 \ \exists N_\varepsilon \in \mathbb{N} \) such that \( \forall n \geq N_\varepsilon \)

\[
\frac{1}{d_n} \# \{ i : \sigma_i^{(n)} \leq \varepsilon \} \leq x(\varepsilon),
\]

where \( \{\sigma_i^{(n)}\}, \ i = 1, \ldots, d_n \) denotes the complete set of the singular values of \( X_n \).

Moreover \( \{X_n\} \) is defined as \emph{sparsely unbounded} if there exists a nonnegative function \( x(s) \) with \( \lim_{s \to 0} x(s) = 0 \) so that \( \forall \varepsilon > 0 \ \exists N_\varepsilon \in \mathbb{N} \) such that \( \forall n \geq N_\varepsilon \)

\[
\frac{1}{d_n} \# \{ i : \sigma_i^{(n)} \geq \frac{1}{\varepsilon} \} \leq x(\varepsilon).
\]
It is worth stressing that the reason for the previous definition is due to the notion of sparsely vanishing Lebesgue measurable functions introduced by Tyrtyshnikov as those functions whose set of zeros has zero Lebesgue measure [30]. In fact, a sequence \( \{X_n\} \) spectrally distributed as a sparsely vanishing function is sparsely vanishing in the sense of Definition 3.1 and a sequence of matrices \( \{X_n\} \) spectrally distributed as a sparsely unbounded function is sparsely unbounded in the sense of Definition 3.1. In Proposition 3.1 we prove the above statements.

**Proposition 3.1.** Let \( \{A_n\}, A_n \in \mathbb{C}^{n \times n} \), be a sequence of matrices spectrally distributed as a sparsely vanishing (sparsely unbounded) function \( f \). Then the sequence \( \{A_n\} \) is sparsely vanishing (sparsely unbounded).

**Proof.** First, we consider the case of a sparsely vanishing function \( f \). For any \( \varepsilon > 0 \) define the nonnegative test function

\[
G_{\varepsilon}(y) = \begin{cases} 
\frac{y + 1}{\varepsilon} & \text{for } -c \leq y \leq 0 \\
1 & \text{for } 0 \leq y \leq \varepsilon \\
\frac{-y + 2}{\varepsilon} & \text{for } \varepsilon \leq y \leq 2\varepsilon \\
0 & \text{otherwise.}
\end{cases}
\]

Now, since

\[
\frac{1}{n} \sum_{i=1}^{n} G_{\varepsilon}(\sigma_{i}(n)) = \frac{1}{n} \left[ \sum_{i \in \{1 \leq |\sigma_{i}(n)| \leq \varepsilon\}} 1 + \sum_{i \in \{1 \leq |\sigma_{i}(n)| \leq 2\varepsilon\}} G_{\varepsilon}(\sigma_{i}(n)) \right] \leq \frac{1}{n} \sum_{i \in \{1 \leq |\sigma_{i}(n)| \leq \varepsilon\}} 1
\]

we find that \( \frac{1}{n} \# \{ j : \sigma_{j}(n) \leq \varepsilon \} \geq \frac{1}{n} \sum_{i=1}^{n} G_{\varepsilon}(\sigma_{i}(n)) \). Moreover,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} G_{\varepsilon}(\sigma_{i}(n)) = \frac{1}{m(K)} \int_{K} G_{\varepsilon}(|f(t)|) \, dt \leq \frac{1}{m(K)} \sum_{x \in K : |f(x)| \leq 2\varepsilon} m(x).
\]

Since \( f \) is sparsely vanishing, then, by definition, \( \lim_{n \to \infty} m\{ x \in K : |f(x)| \leq \eta \} = 0 \), and the thesis directly follows by considering \( \sum_{i=1}^{n} F_{\varepsilon}(\sigma_{i}(n)) \leq \frac{1}{m(K)} \sum_{x \in K : |f(x)| \leq 1} 1 \).

Now, we consider the case of a sparsely unbounded function \( f \). For any \( \varepsilon > 0 \) define the nonnegative test function

\[
F_{\varepsilon}(y) = \begin{cases} 
\frac{y + 1}{\varepsilon} & \text{for } -c \leq y \leq 0 \\
1 & \text{for } 0 \leq y \leq \frac{1}{2\varepsilon} \\
-2\varepsilon y + 2 & \text{for } \frac{1}{2\varepsilon} \leq y \leq \frac{1}{\varepsilon} \\
0 & \text{otherwise.}
\end{cases}
\]

By taking into account the relations below:

\[
\frac{1}{n} \sum_{i=1}^{n} F_{\varepsilon}(\sigma_{i}(n)) = \frac{1}{n} \left[ \sum_{i \in \{1 \leq |\sigma_{i}(n)| \leq \frac{1}{\varepsilon}\}} 1 + \sum_{i \in \{1 \leq |\sigma_{i}(n)| \leq \frac{1}{2\varepsilon}\}} F_{\varepsilon}(\sigma_{i}(n)) \right] \leq \frac{1}{n} \sum_{i \in \{1 \leq |\sigma_{i}(n)| \leq \frac{1}{2\varepsilon}\}} 1
\]

we easily deduce that \( \frac{1}{n} \# \{ j : \sigma_{j}(n) \geq \frac{1}{\varepsilon} \} \geq \frac{1}{n} \sum_{i=1}^{n} F_{\varepsilon}(\sigma_{i}(n)) \). Moreover,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} F_{\varepsilon}(\sigma_{i}(n)) = \frac{1}{m(K)} \int_{K} F_{\varepsilon}(|f(t)|) \, dt \geq \frac{1}{m(K)} \sum_{x \in K : |f(x)| \leq \frac{1}{2\varepsilon}} m(x) = 1 - \frac{1}{m(K)} \sum_{x \in K : |f(x)| > \frac{1}{2\varepsilon}} m(x)
\]

By the inequality \( \frac{1}{n} \# \{ j : \sigma_{j}(n) \geq \frac{1}{\varepsilon} \} = 1 - \frac{1}{n} \# \{ j : \sigma_{j}(n) < \frac{1}{\varepsilon} \} \leq 1 - \frac{1}{n} \sum_{i=1}^{n} F_{\varepsilon}(\sigma_{i}(n)) \) and by recalling that \( f \) sparsely unbounded means that \( \lim_{n \to \infty} m\{ x \in K : |f(x)| \geq \frac{1}{n} \} = 0 \), the thesis now follows by considering \( x(\varepsilon) = 1 - \frac{1}{m(K)} \sum_{x \in K : |f(x)| \geq \frac{1}{2\varepsilon}} m(x) \) in Definition 3.1. \( \square \)
It is worth noticing that essentially the same proof of Proposition 3.1 applies in the case of a sequence of Hermitian matrices with a real-valued function $f$ when considering the eigenvalues instead of the singular values. The only change in the previous proof is in the definition of the test functions $F_n$ and $G_n$: in fact it is enough to take new test functions $\tilde{T}_n = \tilde{T}_n(y)$ that coincide with $T_n(y)$ if the argument $y$ is nonnegative and coincide with $T_n(-y)$ otherwise. Here the symbol “F” or “G” according to the previous notations.

The following result is very useful in practical manipulations in order to give norm bounds from above.

**Lemma 3.1.** Consider a sequence of matrices $\{X_n\}$, with $X_n$ of size $d_n$. The following are equivalent.

- The sequence $\{X_n\}$ is sparsely unbounded.
- There exists a nonnegative function $x(s)$ with $\lim_{s \to 0} x(s) = 0$ so that $\forall \varepsilon > 0 \exists n_\varepsilon \in \mathbb{N}$ such that $\forall n \geq n_\varepsilon$ it holds that $X_n = B_n + L_n$, where $\|B_n\|_2 < \frac{1}{\delta}$ and $\text{rank}(L_n) \leq x(\varepsilon)d_n$.

**Proof.** The result trivially follows by using the singular value decomposition properties of the involved matrices and the singular values interlacing properties [31]. □

The following technical lemmas will be useful for performing the spectral analysis of preconditioned matrices in the next section.

**Lemma 3.2.** Let $\{X_n\}$ and $\{Y_n\}$, $X_n$, $Y_n$ of size $d_n$, be two sparsely unbounded matrix sequences. Then the sequences $\{X_nY_n\}$ and $\{X_n + Y_n\}$ are sparsely unbounded (the latter implies that the notion sparsely unbounded sequence is stable under linear combinations).

**Proof.** Under these assumptions, we can consider the following splittings

\[ X_n = \tilde{B}_n + \tilde{L}_n \]
\[ Y_n = \tilde{B}_n + \tilde{L}_n \]

where $\forall \delta > 0 \exists n_\delta \in \mathbb{N}$ such that $\forall n \geq n_\delta$ it holds that $\|\tilde{B}_n\|_2 < \frac{1}{\delta}$ and $\text{rank}(\tilde{L}_n) \leq \hat{x}(\delta)d_n$ with $\lim_{\delta \to 0} \hat{x}(s) = 0$ and where $\forall \delta > 0 \exists n_\delta \in \mathbb{N}$ such that $\forall n \geq n_\delta$ it holds that $\|\tilde{B}_n\|_2 < \frac{1}{\delta}$ and $\text{rank}(\tilde{L}_n) \leq \hat{x}(\delta)d_n$ with $\lim_{\delta \to 0} \hat{x}(s) = 0$. Therefore, the matrices $X_nY_n$ can be written as

\[ X_nY_n = B_n + L_n \]

with

\[ B_n = \tilde{B}_n\tilde{B}_n, \]
\[ L_n = \tilde{L}_n(\tilde{B}_n + \tilde{L}_n) + \tilde{B}_n\tilde{L}_n, \]

where, for $n$ large enough, we find

\[ \|B_n\|_2 < \frac{1}{\delta}, \]
\[ \text{rank}(L_n) \leq (\hat{x}(\delta) + \hat{x}(\delta))d_n. \]

For the arbitrariness of $\delta$ and $\hat{x}$ the first part of the claimed thesis follows by virtue of Lemma 3.1.

The matrices $X_n + Y_n$ can be written as

\[ X_n + Y_n = \tilde{B}_n + \tilde{L}_n \]

with

\[ \tilde{B}_n = \tilde{B}_n + \tilde{B}_n, \]
\[ \tilde{L}_n = \tilde{L}_n + \tilde{L}_n, \]

where, for $n$ large enough, we find

\[ \|\tilde{B}\|_2 < \frac{1}{\delta} + \frac{1}{\delta} < 2 \left( \min(\delta, \hat{x}(\delta)) \right)^{-1}, \]
\[ \text{rank}(\tilde{L}_n) \leq (\hat{x}(\delta) + \hat{x}(\delta))d_n. \]

For the arbitrariness of $\delta$ and $\hat{x}$ the second part of the claimed thesis follows by Lemma 3.1. □

**Lemma 3.3.** Let $\{X_n\}$ be a sequence of invertible matrices, with $X_n$ of size $d_n$. If the sequence $\{X_n\}$ is sparsely vanishing then the sequence $\{X_n^{-1}\}$ is sparsely unbounded and vice versa.
Proof. The result trivially follows by using the singular value decomposition properties of the involved matrices. \[\square\]

**Lemma 3.4.** Let \(\{X_n\}\) and \(\{Y_n\}\) be two sparsely vanishing matrix sequences of invertible matrices, with \(X_n, Y_n\) of size \(d_n\). Then the sequence \(\{X_nY_n\}\) is sparsely vanishing. This is not true for the sequence \(\{X_n + Y_n\}\), that is, the notion sparsely vanishing sequence is not stable under linear combinations.

**Proof.** Since \(\{X_n\}\) and \(\{Y_n\}\) are both sequences of invertible matrices, from \((X_nY_n)^{-1} = (Y_n)^{-1}(X_n)^{-1}\), the first part trivially follows from Lemma 3.2 by recalling Lemma 3.3. The second part is straightforward by considering \(Y_n = -X_n\), so that \(X_n + Y_n \equiv 0\) is not sparsely vanishing. \[\square\]

**Remark 3.1.** The assumption of invertibility in Lemmas 3.3 and 3.4 can be removed by considering the pseudo-inverse of Moore–Penrose \([32,33]\) instead of the usual inverse matrix.

**Lemma 3.5.** Let \(\{X_n\}\) and \(\{Y_n\}\) be two matrix sequences, with \(X_n, Y_n\) of size \(d_n\). Suppose that the sequence \(\{X_n\}\) is sparsely unbounded and the sequence \(\{Y_n\}\) is clustered at 0 with respect to the singular values. Then both the sequences \(\{X_nY_n\}\) and \(\{X_n + Y_n\}\) are clustered at 0.

**Proof.** Under these assumptions, we have that \(\forall \hat{\varepsilon} > 0\ \exists \eta \in \mathbb{N}\) such that \(\forall n \geq n_{\eta}\) it holds that
\[X_n = B_n + L_n\]
where \(\|B_n\|_2 < \frac{1}{\hat{\varepsilon}}\) and \(\text{rank}(L_n) \leq \chi(\hat{\varepsilon})d_n\) with \(\lim_{n \to \infty} \chi(s) = 0\) and \(\forall \varepsilon > 0\ \exists \eta \in \mathbb{N}\) such that \(\forall n \geq n_{\eta}\) we have
\[X_n = N_n + R_n\]
where \(\|N_n\|_2 \leq \varepsilon\) and \(\text{rank}(L_n) \leq \chi(\varepsilon)d_n\) with \(\lim_{n \to \infty} \chi(s) = 0\). Now, by splitting the matrices as
\[X_nY_n = \tilde{N}_n + \tilde{R}_n\]
with
\[\tilde{N}_n = B_nN_n, \quad \tilde{R}_n = B_nR_n + L_n(N_n + R_n),\]
where
\[\|\tilde{N}_n\|_2 \leq \frac{\varepsilon}{\hat{\varepsilon}}, \quad \text{rank}(\tilde{R}_n) \leq (\chi(\varepsilon) + \chi(\hat{\varepsilon}))d_n\]
and for the arbitrariness of \(\hat{\varepsilon}\) and \(\varepsilon\), by choosing \(\hat{\varepsilon} = \sqrt{\varepsilon}\), the desired result plainly follows. The case \(\{Y_nX_n\}\) can be proved in the same manner. \[\square\]

**Lemma 3.6.** Consider a sequence \(\{A_n\}\), where \(A_n\) is of size \(d_n\). Then the following are equivalent.

- There exists a sequence \(\{D_n\}\) so that \(\|A_n - D_n\|_2^2 = o(d_n)\) and \(\text{rank}(D_n) = o(d_n)\).
- There exists a sequence \(\{D_n\}\) so that \(\forall p \in [1, +\infty]\) it holds \(\|A_n - D_n\|_p^p = o(d_n)\), \(\text{rank}(D_n) = o(d_n)\). || \(\|\cdot\|_p\) denoting the Schatten p-norm of matrices \([34]\).
- There exists a function \(\chi(s)\) such that \(\lim_{n \to \infty} \chi(s) = 0\) so that \(\forall \varepsilon > 0\ \exists n_{\varepsilon} \in \mathbb{N}\) such that \(\forall n \geq n_{\varepsilon}\) it holds \(A_n = N_n + R_n\), with \(\|N_n\|_2 \leq \varepsilon\) and \(\text{rank}(R_n) \leq \chi(\varepsilon)d_n\).
- The sequence \(\{A_n\}\) is clustered at zero (refer to Definition 2.2).
- The sequence \(\{A_n\}\) is spectrally distributed as the identically null function (refer to Definition 2.1).  

**Proof.** It is a direct check by making a clever use of the singular value decomposition \([31]\). \[\square\]

**Lemma 3.7.** Consider two sequences \(\{A_n\}\) and \(\{B_n\}\), where \(A_n, B_n\) are of size \(d_n\). If there exists a sequence \(\{D_n\}\) so that \(\|A_n - B_n - D_n\|_2^2 = o(d_n)\) and \(\text{rank}(D_n) = o(d_n)\), then the sequence \(\{A_n - B_n\}\) is spectrally distributed as the identically null function (in the sense of Definition 2.1) and the sequences \(\{A_n\}\) and \(\{B_n\}\) are equally distributed (in the sense of the last part of Definition 2.1). In addition, if one of the sequences is spectrally distributed as a function then the other sequence possesses the same distribution.

**Proof.** By the equivalence Lemma 3.6 we get that \(\{A_n - B_n\} \sim_{o} 0\). The equal distribution of the sequences \(\{A_n\}\) and \(\{B_n\}\) was proved by Tsybakov \([26]\). Lastly, if one of the sequences is spectrally distributed as a function then, by definition of equal distribution, it is easy to recognize that the other sequence possesses the same distribution. \[\square\]

**Theorem 3.1.** Let \(\{X_n\}\) and \(\{P_n\}\) be two sequences of matrices, with \(X_n, P_n\) of size \(d_n\). Let \(\{I_n\}\) be the sequence of identity matrices of size \(d_n\). Suppose that the sequence \(\{X_n\}\) is sparsely unbounded, the matrices \(P_n\) are all invertible and the sequence \(\{P_n^{-1}X_n - I_n\}\) is clustered at 0. Then \(\{X_n - P_n\} \sim_{o} 0\) and the sequences \(\{P_n\}\) and \(\{X_n\}\) are equally distributed. In addition, if the sequence \(\{X_n\}\) is distributed as a function then the sequence \(\{P_n\}\) has the same distribution.

Finally, if \(\{X_n - P_n\} \sim_{o} 0\) that is \(\{X_n - P_n\}\) is clustered at 0, then \(\{P_n^{-1}X_n - I_n\}\) is clustered at 0, under the condition that \(\{P_n^{-1}\}\) is sparsely vanishing, that is \(\{P_n\}\) is sparsely unbounded.
Lemma 3.6 implies Lemma 3.7.

Lemma 3.5 tells us that the set of matrices \( \{ X_n \} \) is clustered at zero. For the last part, we just observe that \( P_n^{-1}X_n - I_n = P_n^{-1}(X_n - P_n) \) so that Lemma 3.5 implies \( P_n^{-1}X_n - I_n \sim 0 \) if \( \{ X_n - P_n \} \sim 0 \) and \( P_n^{-1} \) is sparsely unbounded (which is the same as \( P_n \) is sparsely vanishing given the invertibility of each \( P_n \) and thanks to Lemma 3.3).

Remark 3.2. Lemma 3.2 tells us that the set of sparsely unbounded sequences forms an algebra (that is closed under linear combinations and products). On the other hand, Lemma 3.5 can be read in an abstract way, by saying that the set of sequences which are clustered at zero forms a two-sided ideal in the algebra of sparsely unbounded sequences.

Remark 3.3. Theorem 3.1 has a “philosophical” meaning. If we think of the matrices \( P_n \) as preconditioners, then Theorem 3.1 states that a good preconditioning sequence \( \{ P_n \} \) inherits from the original sequence \( \{ X_n \} \) the distribution, if any. Moreover if the sequence \( \{ X_n \} \) is sparsely unbounded (sparsely vanishing) then the same is true for the sequence \( \{ P_n \} \).

Remark 3.4. The sparsely unboundedness assumption of \( \{ X_n \} \) is necessary and cannot be removed as far as we are concerned with Theorem 3.1. For instance, take \( X_n = (n + 1)I_n \) and \( P_n = nI_n \). Then the sequence \( \{ P_n^{-1}X_n - I_n \} = \{ \frac{1}{n+1} \} \) is clustered at 0, but \( \{ X_n - P_n \} = \{ I_n \} \) is not. However \( \{ X_n \} \) and \( \{ P_n \} \) have the same distribution function, since they are both distributed as the constant function \( \infty \).

Theorem 3.2. Let \( \{ X_n \} \), \( \{ Y_n \} \) and \( \{ P_n \} \) be three sequences of matrices, with \( X_n, Y_n, P_n \) of size \( d_n \) and \( P_n \) invertible for any \( n \). Let \( \{ I_n \} \) be the sequence of identity matrices of size \( d_n \). Suppose that

1. the sequence \( \{ X_n \} \) is sparsely vanishing,
2. the sequence \( \{ X_n - Y_n \} \) is clustered at 0,
3. the sequence \( \{ P_n^{-1}X_n - I_n \} \) is clustered at 0.

Then the sequence \( \{ P_n^{-1}Y_n - I_n \} \) is clustered at 0.

Proof. The matrices \( P_n^{-1}Y_n - I_n \) can clearly be split as

\[
P_n^{-1}Y_n - I_n = (P_n^{-1}X_n - I_n) + P_n^{-1}(Y_n - X_n),
\]

(5)
where the sequence \( \{ P_n^{-1}X_n - I_n \} \) is clustered at 0 by assumption 3. Moreover the sequence \( \{ P_n \} \) is sparsely vanishing since the sequence \( \{ X_n \} \) is sparsely vanishing (see Remark 3.3). Therefore the application of Lemmas 3.3 and 3.5 proves that the sequence \( \{ P_n^{-1}(Y_n - X_n) \} \) is clustered at 0. As a final statement, in the light of Eq. (5), the sequence \( \{ P_n^{-1}Y_n - I_n \} \) is expressed as the sum of two matrix sequences that are clustered at 0, so that the proof is concluded, by invoking the first claim of Proposition 2.1 with \( \theta = 0 \). \( \square \)

4. Singular value distribution of g-circulants and g-Toeplitz sequences

Let \( f \) be a Lebesgue integrable function defined on \( (-\pi, \pi)^d \) and taking values in \( \mathcal{M}_{pq} \), for given positive integers \( p \) and \( q \). Then, for \( d \)-indices \( r = (r_1, \ldots, r_d) \), \( s = (s_1, \ldots, s_d) \), \( n = (n_1, \ldots, n_d) \), \( e = (1, \ldots, 1) \), \( \beta = (0, \ldots, 0) \), the Toeplitz matrix \( T_n(f) \) of size \( pn \times \hat{qn} \) is defined as follows: \( T_n(f) = [ E_{\alpha r s t}^{p q n }]_{\alpha s t \in \mathbb{Z}^d} \), where \( E_{\alpha r s t}^{p q n } \) are the Fourier coefficients of \( f \)
defined by equation

\[
E_{\alpha r s t}^{p q n } = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} f(t_1, \ldots, t_d) e^{-i(\alpha r_1 t_1 + \cdots + \alpha s_d t_d)} dt_1 \cdots dt_d, \quad \alpha^2 = -1,
\]

for integers \( \alpha \). Since \( f \) is a matrix-valued function of \( d \)-variables whose component functions are all integrable, then the \( (k_1, \ldots, k_d) \)-th Fourier coefficient is considered to be the matrix whose \((u, v)\)-th entry is the \( (k_1, \ldots, k_d) \)-th Fourier coefficient of the function \( (f(t_1, \ldots, t_d)_{u,v}) \).

According to this multi-index block notation we can define general multilevel block g-Toeplitz and g-circulants. Of course, in this multidimensional setting, \( g \) denotes a \( d \)-dimensional vector of nonnegative integers that is \( g = (g_1, \ldots, g_d) \).

In that case \( A_n = T_{n,g} = [a_{\alpha r s t}^{p q n }]_{r s t \in \mathbb{Z}^d} \), where \( a_{\alpha r s t}^{p q n } = E_{\alpha r s t}^{p q n } \) is called g-circulant if \( A_n = C_{n,g} = [a_{(r - g \circ s) \alpha}^{p q n }]_{r s t \in \mathbb{Z}^d} \), where\( (r - g \circ s) \mod n = ((r_1 - g_1 s_1) \mod n_1, (r_2 - g_2 s_2) \mod n_2, \ldots, (r_d - g_d s_d) \mod n_d) \).

4.1. The singular value distribution result for g-Toeplitz sequences

We consider the general multilevel case, where \( f \) is allowed to be both Lebesgue integrable over \( Q^d \) and matrix-valued, \( Q = (-\pi, \pi) \). We have\( \{ T_n(f) \} \sim_{o} (\theta_t, Q^d \times [0,1]^d) \), (6)

where

\[
\theta_t(x, t) = \begin{cases} \sqrt{|f(t)|^2(x)} & \text{for } t \in \left[0, \frac{1}{g}\right], \\ 0 & \text{for } t \in \left(\frac{1}{g}, 1\right) \end{cases}
\]

with

\[
|f(t)|^2(x) = \frac{1}{g} \sum_{j=0}^{g-1} |f\left(x + \frac{2\pi j}{g}\right)|^2,
\]

and where all the arguments are modulus \( 2\pi \) and all the operations are intended componentwise, that is \( t \in \left[0, \frac{1}{g}\right] \) means that \( t_k \in \left[0, \frac{1}{g}\right], k = 1, \ldots, d \), \( t \in \left(\frac{1}{g}, 1\right) \) means that \( t_k \in \left(\frac{1}{g}, 1\right], k = 1, \ldots, d \), the writing \( \frac{x + 2\pi j}{g} \) defines the \( d \)-dimensional vector whose \( k \)-th component is \( \frac{x_k + 2\pi j_k}{g_k} \), \( k = 1, \ldots, d \), and \( g = g_1 g_2 \cdots g_d \). Moreover, if the vector \( g \) is degenerate, namely there exists an index \( s \in \{1, \ldots, d\} \) for which \( g_s = 0 \), then the function \( \sqrt{|f(t)|^2(x)} \) becomes identically zero so that\( \{ T_n,g \} \sim_{o} (0, G) \)

for every admissible set \( G \). For some concrete examples of g-circulant and g-Toeplitz sequences and related spectra, where some of the entries of \( g \) vanish, see [35]. Interestingly enough, if \( g \) is the vector of all ones, that is we are in the standard Toeplitz multilevel context, then \( T_{n,g} = T_n(f) \), \( \sqrt{|f(t)|^2(x)} \) reduces to \( |f(x)| \), and the variable \( t \in [0,1]^d \) becomes useless so that\( \{ T_n(f) \} \sim_{o} (f, Q^d \times [0,1]^d) \)
which is the same as the classical Szegö–Tyrtyshnikov–Tilli result [23,24]

\[ T_a(f) \sim_\sigma f^\ast, Q^\dagger. \]

We finally mention that the technique for obtaining formula (6), as in locally Toeplitz setting [36,37], strongly relies on the notion of the approximating class of sequences [27] which was aimed to develop a basic approximation theory, when the spectral distribution of matrix sequences is considered.

4.2. The singular value distribution result for \( g \)-circulant sequences

In the following, let \( \gcd(a, b) \) denote the great common divisor of the integer numbers \( a \) and \( b \). Following the analysis in [12], for \( g \) fixed vector and \( n \) increasing sequence of vectors we do not find a joint distribution. Assuming \( \{C_n\} \sim_\sigma (h, Q^\dagger) \) with \( \{C_n\} \) standard sequence of multilevel circulants (that is \( g \)-circulants where \( g \) is the vectors of all ones), and assuming that the sequence \( n \) is chosen so that \( \gamma_i = \gcd(n_i, g_i), i = 1, \ldots, d \), are \( d \) fixed numbers, we find

\[ \{C_{n,g}\} \sim_\sigma (\eta_n, Q^d \times [0, 1]^d), \]

where

\[ \eta_n(x, t) = \begin{cases} \sqrt{|h|^3(x)} & \text{for } t \in \left[ 0, \frac{1}{\gamma} \right], \\ 0 & \text{for } t \in \left( \frac{1}{\gamma}, 1 \right]. \end{cases} \]

(10)

with \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_d) \),

\[ |h|^3(x) = \hat{\gamma}|h|^2(x) = \sum_{j=0}^{ve} |h|^2 \left( \frac{x + 2\pi j}{\gamma} \right), \]

and \( \hat{\gamma} = \gamma_1\gamma_2 \cdots \gamma_d \).

5. Preconditioning of \( g \)-Toeplitz sequences via \( g \)-circulant sequences

We start by analyzing the possibility of a standard preconditioning in the light of the distribution results and of the analysis of Section 3. Then we consider the preconditioning in a regularizing context.

5.1. Consequences of the distribution results on preconditioning of \( g \)-Toeplitz sequences

We study the possibility of a standard preconditioning taking into consideration the distribution results and of the analysis of Section 3.

First of all Theorem 3.1 tells one that \( \{P_n\} \) is a good preconditioning sequence for \( \{X_n\} \) (that is \( \{P_n^{-1}X_n - I_n\} \sim_\sigma 0 \)) if and only if \( \{X_n - P_n\} \sim_\sigma 0 \) and \( \{P_n\} \) sparsely vanishing, with the matrices \( P_n \) all invertible. The consequences below are of paramount importance:

- The vector \( g \) has to be strictly positive; if not the original problem \( T_{n,g}x = b \) is substantially ill-posed since \( T_{n,g} \sim_\sigma 0 \) and in addition \( C_{n,g} \) is singular and indeed \( \{C_{n,g}\} \sim_\sigma 0 \) which violates the crucial condition of Theorem 3.1 that \( \{P_n\} \) is sparsely vanishing with \( P_n = C_{n,g} \).
- Even in the case that \( g \) is strictly positive, relations (6)–(8) imply that \( \{X_n\} \) with \( X_0 = T_{n,g} \) is sparsely vanishing if and only if \( f \) is sparsely vanishing and \( g_i = 1 \) (or more generally \( g_i = \pm 1 \), \( i = 1, \ldots, d \)). In other words, again by Theorem 3.1, a good preconditioning can be achieved only in the standard case of multilevel Toeplitz sequences and in fact the latter is a case widely studied in the literature [13–15] (for \( d = 1 \) also with strong clustering when \( f \) is continuous [15], while for \( d > 1 \) the clustering is necessarily weak due to the computational barrier proven in [38]).
- In any case the condition required by Theorem 3.1 that the sequences \( \{X_n\}, \{P_n\} \), with \( X_n = T_{n,g}, P_n = C_{n,g} \), share the same distribution symbol is quite tricky. By comparing (6)–(11), the latter is possible only for the case where \( g_i = \gcd(n_i, g_i), i = 1, \ldots, d \), and we have to choose \( h = \frac{1}{\gamma} f \).

In conclusion, a good preconditioning can be reached only in the standard multilevel Toeplitz setting. However, by looking at the preconditioning in a different sense, something useful can be said.

5.2. Regularizing preconditioning

Suppose that \( \{X_n\} \) is a sequence of matrices with \( X_n \) of size \( d_n \) and there exists a sequence of subspaces \( \{\delta_n\} \) of size \( r_n \) being the integer part of \( cd_n, c \in (0, 1) \) for which \( \forall \epsilon > 0, \exists r_n \) and

\[ \|X_nv\| \leq \epsilon \|v\|, \quad \forall v \in \delta_n, \forall n \geq n_\epsilon. \]
This situation naturally arises when \( \{X_n\} \sim_{\sigma} (\theta, G) \) with \( \theta \) vanishing on \( \hat{G} \subset G \) with \( \frac{m(\hat{G})}{m(G)} = c, m(\cdot) \) being the Lebesgue measure and \( |\theta| > 0 \) almost everywhere in the complement \( G \setminus \hat{G} \). Under such conditions we look for a preconditioning \( \{J_n\} \) already in inverse form such that
\[
\|J_nX_nv\| \leq \epsilon \|v\|, \quad \forall v \in \delta_n, \forall n \geq \tilde{n}_\epsilon,
\]
and
\[
\|J_nX_nv - v\| \leq \epsilon \|v\|, \quad \forall v \in \delta_n^\perp, \forall n \geq \tilde{n}_\epsilon.
\]
In other words \( J_nX_n \) when restricted to \( \delta_n \) is close to the null matrix, while it is close to the identity matrix in the orthogonal complement.

Let now consider a \( d_n \times d_n \) matrix \( M_n \) such that its first \( d_n - r_n \) columns are a basis for the subspace \( \delta_n^\perp \) and the other \( r_n \) columns are a basis for \( \delta_n \). These conditions allow us to write that \( J_nX_n \) is an \( \epsilon \)-perturbation of
\[
M_n \begin{bmatrix} I_{d_n-r_n} & 0 \\ 0 & 0_{r_n} \end{bmatrix} M_n^{-1},
\] (12)
where \( I_{d_n-r_n} \) is the identity matrix of size \( d_n - r_n \) and \( 0_{r_n} \) is the null matrix of order \( r_n \).

We mention that, in general, if the subspace \( \delta_n \) is the space where the noise usually dominates (which in general is related to high frequencies in inverse problems) and moreover its dimension \( r_n \) is a parameter which can be tuned (hence, the dimension of the subspace where \( J_nX_n \) approximates the null matrix can be enlarged or made smaller), then the sequence \( J_n \) satisfying (12) is called a regularizing preconditioner for \( X_n \) [39].

In the next Section 5.4, we will show that, if \( X_n \) is a sequence of \( g \)-Toeplitz matrices \( T_{m,g} \), and \( J_m \) is the corresponding sequence of \( g \)-circulant matrices \( C_{m,g} \), then the above factorization arises with the very special case in which \( M_n = I_n \) and \( r_n = d_n - \lceil n/g \rceil \), being \( d_n = n \). This will help us to consider \( g \)-circulant matrices as useful preconditioners for \( g \)-Toeplitz matrices. Unfortunately, the crucial condition that the subspace \( \delta_n \) is the space where the noise usually dominates is not satisfied in general. As we will show in the numerical section, the \( g \)-circulant can be considered as regularizing preconditioner for \( g \)-Toeplitz sequences provided that a regularizing technique (depending on an appropriate and tunable regularization parameter) is applied to satisfy this crucial condition.

5.3. Some preparatory tools

In the following, in order to compact the heavy notation, we often denote by \( (n, g) \) the greatest common divisor of \( n \) and \( g \), i.e., \( (n, g) = \gcd(n, g) \) (both notations will be used), and the integer numbers \( n_g \) and \( \hat{g} \) are defined respectively as
\[
n_g = \frac{n}{(n, g)} \quad \text{and} \quad \hat{g} = \frac{g}{(n, g)}.
\]
Since the notations can become quite heavy, for the sake of simplicity and at the beginning, we start with the case \( d = p = q = 1 \). Several generalizations are given in Section 5.5. We observe that also the case of nonpositive \( g \) can be taken into consideration and can be reduced to the case of a nonnegative \( g \). In fact, the role of circulants will be played by \((-1\)-circulant matrices (also called anti-circulants or skew-circulants), [1]: as for the circulants, \((-1\)-circulants form a commutative algebra simultaneously diagonalized by another unitary transform that can be written as the product of the Fourier matrix and a diagonal unitary matrix.

By direct simple computation, for generic \( n \) and \( g \) one immediately finds that \( C_{n,g} = C_n \tilde{Z}_{n,g} \), where
\[
Z_{n,g} = \left[ \delta_{r-g,0} \right]_{r,s=0}^{n-1}, \quad \delta_k = \begin{cases} 1 & \text{if } k \equiv 0 \pmod{n} \\ 0 & \text{otherwise.} \end{cases} \]
(13)
The following preparatory results are straightforward. The detailed proofs are reported in [35]; see also [1].

Lemma 5.1 ([35]). Let \( n \) be any integer greater than 2 then
\[
Z_{n,g} = \left[ \tilde{Z}_{n,g} \tilde{Z}_{n,g} \cdots \tilde{Z}_{n,g} \right]_{(n,g)} \text{ times},
\]
where \( \tilde{Z}_{n,g} \) is the matrix defined in (13) and \( \tilde{Z}_{n,g} \in \mathbb{C}^{n \times n_g} \) is the submatrix of \( Z_{n,g} \) obtained by considering only its first \( n_g \) columns, that is
\[
\tilde{Z}_{n,g} = Z_{n,g} \begin{bmatrix} I_{n_g} \\ 0 \end{bmatrix}.
\]
Moreover
\[
\tilde{Z}_{n,g} = \tilde{Z}_{n,(n,g)} Z_{n_g,\hat{g}},
\]
where \( Z_{n_g,\hat{g}} \) is the matrix defined in (13) of dimension \( n_g \times n_g \). Therefore
\[
Z_{n_g,\hat{g}} = \left[ \delta_{r-h,0} \right]_{r,s=0}^{n_g-1}, \quad \delta_k = \begin{cases} 1 & \text{if } k \equiv 0 \pmod{n_g} \\ 0 & \text{otherwise.} \end{cases} \]
(15)
Finally, if \( g \geq n \) then \( Z_{n, g} = Z_{n, g^*} \), where \( g^* = g \mod n \) and \( Z_{n, g} \) is defined in (13), so that
\[
C_{n, g} = C_n Z_{n, g} = C_n Z_{n, g^*} = C_n g^*
\]
and
\[
C_n = F_n D_n F_n^* Z_{n, g}.
\]
\[
F_n = \frac{1}{\sqrt{n}} \left[ e^{\frac{2\pi i k}{n}} \right]_{j=0}^{n-1}, \quad \text{Fourier matrix,}
\]
\[
D_n = \text{diag}(\sqrt{n} e_n^g), \quad e_n^g = [a_0, a_1, \ldots, a_{n-1}]^T, \quad \text{first column of the matrix } C_n.
\]

**Lemma 5.2** ([35]). Let \( F_n \) be the Fourier matrix of order \( n \) defined in (16) and let \( \tilde{Z}_{n, g} \in \mathbb{C}^{n \times n g} \) be the matrix represented in (14). Then
\[
F_n \tilde{Z}_{n, g} = \frac{1}{\sqrt{(n, g)}} I_{n, g} F_{n, g} Z_{n, g^*},
\]
where \( I_{n, g} \in \mathbb{C}^{n \times n g} \) and
\[
I_{n, g} = \left[ \begin{array}{c} I_{n, g} \\ \vdots \\ I_{n, g} \end{array} \right] \quad \text{(n, g) times},
\]
with \( I_{n, g} \) being the identity matrix of size \( n \) and \( Z_{n, g^*} \) as in (15). Therefore \( \tilde{Z}_{n, g} \) is the matrix \( Z_{n, g^*} \) by considering only the first \( \mu_g \) columns, then \( 1 \leq (n, g) \leq \mu_g \leq n \) and
\[
\tilde{Z}_{n, g} = \left[ \begin{array}{c} I_{n, g} \\ 0 \end{array} \right].
\]

**Remark 1.** In Lemma 5.2, if \( (n, g) = g \), we have \( n_g = \frac{n}{(n, g)} = \frac{n}{g} \) and \( \tilde{g} = \frac{g}{(n, g)} = 1 \); so the matrix \( Z_{n_g, \tilde{g}} = Z_{n_g, 1} \), appearing in (17), is the identity matrix of dimension \( \frac{n}{g} \times \frac{n}{g} \). The relation (17) becomes
\[
F_n \tilde{Z}_{n, g} = \frac{1}{\sqrt{g}} I_{n, g} F_{n, g}.
\]

**Remark 2.** If \( (n, g) = 1 \), Lemma 5.2 is trivial, because \( n_g = \frac{n}{(n, g)} = n \), \( \tilde{g} = \frac{g}{(n, g)} = g \), and so \( \tilde{Z}_{n, g} = Z_{n, g} \). The relation (17) becomes
\[
F_n \tilde{Z}_{n, g} = F_n Z_{n, g} = I_{n, g} F_{n, g} Z_{n, g^*} = F_n Z_{n, g^*},
\]
since the matrix \( I_{n, g} \) reduces by its definition to the identity matrix of order \( n \).

**Remark 3.** Lemma 5.2 is true also if, instead of \( F_n \) and \( F_{n, g} \), we put \( F_n^* \) and \( F_{n, g}^* \), respectively, because \( F_n^* = \overline{F_n} \). In fact there is no transposition, but only conjugation.

5.4. The analysis of regularizing preconditioners when \( p = q = d = 1 \) and \( n \) chosen s.t. \( \gcd(n, g) = 1 \)

According to the very concise analysis in Section 5.2, we will prove that a proper choice of the matrix sequence \( \{C_{n, g}\} \) leads to a preconditioning scheme for the sequence \( \{T_{n, g}\} \), such that relation (12) holds, with \( M_n = I_n \), at least when the entries of \( T_{n, g} \) come from the Fourier coefficients of a sparsely vanishing function \( f \). In other words, only the first condition concerning regularizing preconditioning sequences is satisfied.

**Theorem 5.1.** Let \( \{T_{n, g}\} \) be a sequence of \( g \)-Toeplitz matrices generated by a sparsely vanishing function \( f \in L^1(Q) \), with \( Q = (-\pi, \pi) \) and let \( C_n \) be the Frobenius distance minimizer in the standard circulant algebra of the classical Toeplitz matrix \( T_n(f) \). If \( \gcd(n, g) = 1 \), then the sequence \( \{C_{n, g}^{-1}\} \), where \( \{C_{n, g}\} = \{C_n Z_{n, g}\} \), with \( Z_{n, g} \) defined in (13), is a sequence of preconditioners for \( \{T_{n, g}\} \) such that
In 1538 we have that Lemma 3.5. By direct computation, for Lemma 5.2 (13) we obtain vanishing (see [f]).

\[ C_{n,g}^{-1} = \begin{bmatrix} l_{12} & 0 \\ 0 & 0_{n-mg} \end{bmatrix} + V_{n,g}, \]

where \( \mu_k = \left\lfloor \frac{n}{k} \right\rfloor \) and \(|V_{n,g}| \sim 0\).

**Proof.** By direct computation, for \( n \) and \( g \) generic it is simple to verify that

\[ T_{n,g} = \left[ \widehat{T}_{n,g} | T_{n,g} \right] = \left[ T_{n,g} \right] \widehat{T}_{n,g}, \]

\[ = T_{n,g} \left[ \widehat{T}_{n,g} \ | \ 0 \right] + \left[ 0 \ | \ T_{n,g} \right] \right]. \]

For (a constructive proof of relation (18) see [12, pag. 12]). We observe that, since \( \gcd(n,g) = 1 \), \( Z_{n,g} \) is a permutation matrix (see Lemma 5.2), and \( \widehat{Z}_{n,g} \) in (19) is the matrix composed by the first \( n \) columns of \( Z_{n,g} \), \( T_{n,g} \in \mathbb{C}^{n \times (n-mg)} \) is the matrix composed by the \( n - \mu_k \) last columns of \( T_{n,g} \), and \( \widehat{Z}_{n,g} \) is the matrix

\[ \widehat{Z}_{n,g} = \{\delta_{1-gs}\}_{r,s}, \]

\[ r = 0, \ldots, n - 1, \quad s = 0, \ldots, \mu_k - 1, \quad \text{being} \quad \delta_{k} = \begin{cases} 1 & \text{if} \ k \equiv 0 \pmod{n}, \\ 0 & \text{otherwise}, \end{cases} \]

(19)

We now consider the product \( C_{n,g}^{-1}T_{n,g} \) from (18) and since \( Z_{n,g}^{-1}T_{n,g} = I_n \) we have that

\[ C_{n,g}^{-1}T_{n,g} = Z_{n,g}^{-1}C^{-1}T_{n,g} \] and, since \( \{C_{n,g}^{-1}T_{n,g}\} \sim 1 \) or more precisely if \( \{C_{n,g}^{-1}T_{n,g} - I_n\} \sim 0 \), i.e.,

\[ C_{n,g}^{-1}T_{n,g} = I_n + E_{n}, \quad \text{with} \quad \{E_{n}\} \sim 0, \]

we obtain

\[ C_{n,g}^{-1}T_{n,g} = Z_{n,g}^{-1}C^{-1}T_{n,g} + C_{n,g}^{-1}0_{|T_{n,g}|} \]

\[ = Z_{n,g}^{-1}[I_n + E_{n}][Z_{n,g}^{-1}]0_{|T_{n,g}|} + C_{n,g}^{-1}0_{|T_{n,g}|} \]

\[ = Z_{n,g}^{-1}[Z_{n,g}^{-1}]0_{|T_{n,g}|} + Z_{n,g}^{-1}[E_{n}][Z_{n,g}^{-1}]0_{|T_{n,g}|} + C_{n,g}^{-1}0_{|T_{n,g}|} \]

\[ = \left[ \begin{array}{ccc} \frac{I_{n}}{1} & \frac{0}{0} & \frac{0_{n-mg}}{0} \end{array} \right] + Z_{n,g}^{-1}[E_{n}][Z_{n,g}^{-1}]0_{|T_{n,g}|} + C_{n,g}^{-1}0_{|T_{n,g}|}. \]

From Proposition 2.1, since \( \|Z_{n,g}^{-1}\| = 1 \) and \( \|Z_{n,g}^{-1}\| = 1 \) (indeed the first is a permutation matrix and the second is an “incomplete” permutation matrix), and since \( \{E_{n}\} \sim 0 \), we infer that \( \|Z_{n,g}^{-1}[E_{n}][Z_{n,g}^{-1}]0_{|T_{n,g}|}\| \sim 0 \). Moreover, since \( \{C_{n,g}\} \sim \sigma(f, Q) \) with \( f \) sparsely vanishing, from Proposition 3.1 and Lemma 3.3 we have that \( \{C_{n,g}^{-1}\} \) is sparsely unbounded.

Finally, since in [12, Section 4.2.2] it was shown that \( \{0_{|T_{n,g}|}\} \sim 0 \), from Lemma 3.5, we deduce that \( \{C_{n,g}^{-1}0_{|T_{n,g}|}\} \sim 0 \) and the proof is concluded by considering \( V_{n,g} = Z_{n,g}^{-1}E_{n}[Z_{n,g}^{-1}]0_{|T_{n,g}|} + C_{n,g}^{-1}0_{|T_{n,g}|}. \)

**Remark 5.4.** In Theorem 5.1 any preconditioning sequence \( \{C_{n}\} \) for which \( \{C_{n,g}^{-1}T_{n,g} - I_n\} \sim 0 \) will lead to the same thesis. In other words, the choice of the Frobenius optimal preconditioners is just a possible example.

5.5. **Comments on the generalization to the multilevel case** \( d > 1 \)

With all the constraints of Section 5.4 we can have \( d > 1 \), that is \( n = (n_1, \ldots, n_d) \) sequence of integer positive vectors with \( \gcd(n_i, g) = 1, i = 1, \ldots, d, \) so that \( Z_{n,g} \) is still a permutation matrix. The proof reported in Section 5.4 is identical with the only observation that the cluster of \( \{C_{n,g}^{-1}T_{n,g} - I_n\} \) is weak and not strong, due to the computational barrier proven in [38]. More precisely, under the assumption of positivity and continuity of \( f \), by using the Korovkin theory [15] and the Tony Chan preconditioners, we find that the number of outliers of \( \{C_{n,g}^{-1}T_{n,g} - I_n\} \) grows asymptotically as \( \hat{n} \left( \sum_{j=1}^{d} n_j \right) \), \( \hat{n} = \prod_{j=1}^{d} n_j \). Moreover the weak clustering can be achieved by using the mild assumption that \( f \) is only Lebesgue integrable and sparsely vanishing (see [40]).
Furthermore, by following the approach in [16], nothing changes if we assume that the multilevel setting is accompanied by the block setting, i.e. when each entry of the $g$-Toeplitz matrix is a $p \times q$ matrix with complex entries, with $p + q \geq 3$ (somehow the only condition is the recourse to the Moore–Penrose inverse when $p \neq q$).

A bit trickier is the case where the assumption $(n_i, g_i) = 1, i = 1, \ldots , d$, is dropped. In that case $C_{n,g} = C_nZ_{n,g}$ is inherently singular due to the singularity of $Z_{n,g}$ whose rank is $n_g$ with $\mu_g \leq n_g < n$, $\mu_g = \left[ \frac{n}{g} \right]$ (see Lemma 5.2, where all the objects $n, g, \mu_g, n_g, (n, g)$ are $d$-dimensional vectors of positive integers and the inequalities are componentwise).

In this case a good preconditioner already in inverse form is

$$J_n = Z_{n,g}^{-1}C_{n,g}^{-1}$$

with $C_{n,g}$ the usual Tony Chan preconditioner (refer to Section 5.2). Since $\mu_g \leq n_g < n$ (because $1 < (n, g) \leq g$) by Lemma 5.2 we find

$$Z_{n,g}^{-1}C_{n,g}^{-1} = \begin{bmatrix} I_{n_g} & 0 \\ 0 & 0 \end{bmatrix}.$$ 

As a consequence the proof given in Theorem 5.1 is the same and the final result is identical: for the sake of completeness we only observe that the term $C_{n,g}^{-1}C_{n,g}^{-1}$ is always replaced by $Z_{n,g}^{-1}C_{n,g}^{-1}$ so that $\{0|\sigma_n]\} \sim_\sigma 0$ because $\{C_n\} \sim_{\sigma}(f, Q^k)$ with $f$ sparsely vanishing and $\{0|\sigma_n]\} \sim_\sigma 0$ (see Proposition 3.1, Lemmas 3.3 and 3.5 and again [12, Section 4.2.2] where it is shown that $\{0|\sigma_n]\} \sim_\sigma (0, Q)$), and finally $\{Z_{n,g}^{-1}C_{n,g}^{-1}|0|\sigma_n]\} \sim_\sigma 0$ because of Proposition 2.1, where $Z_{n,g}$ plays the role of $Q_n$ and $C_{n,g}^{-1}|0|\sigma_n]\}$ plays the role of $B_n$.

Finally we observe that we have emphasized the role of the Frobenius optimal preconditioner proposed by Tony Chan, for which a very general and rich clustering analysis is available thanks to the Korovkin theory. However, any other alternative and successful preconditioner for standard Toeplitz structures can be employed thanks to Theorem 3.2, which states a kind of useful transitive property.

6. Numerical experiments

Aimed at providing numerical evidence to the theoretical results of the previous section, now we analyze

(i) the distribution of the singular values of $g$-Toeplitz matrices and related $g$-circulant preconditioned matrices (Section 6.1), and

(ii) the effectiveness of the $g$-circulant preconditioning for the solution of the corresponding $g$-Toeplitz linear system $Ax = b$ (Section 6.2).

We remark that, in order to lighten the notation in this subsection, the $g$-Toeplitz systems matrix will be simply denoted as $A$, and its $g$-circulant approximations as $P$. Here $d = p = q = 1$ (i.e. the classical 1D linear systems), whereas the dimension $n$ of the matrices and the value of the parameter $g$ will be explicitly reported in any considered case.

In these numerical experiments, we consider six well-known test cases, most of them firstly used in pioneering works by G. Strang, T. Chan and E. Tyrtyshnikov for the classical Toeplitz preconditioning (i.e., for $g = 1$). For each of any considered test, we report the elements of the first column $(a_0, a_1, \ldots , a_{n-1})^T$ and some properties of the basic symmetric Toeplitz matrix $T_n = [[a_{r-s}]]_{r,s=1}^n$.

- Test 1 $a_k = (k + 1)^{-1}$.
  - Strictly positive non-Wiener generating function, well-conditioned for $g = 1$ (only) [41,14].
- Test 2 $a_k = (k + 1)^{-2}$.
  - Strictly positive Wiener generating function, well-conditioned for $g = 1$ (only) [41,14].
- Test 3 $(a_0, a_1, \ldots , a_{n-1})^T = (2, -1, 0, \ldots , 0)^T$.
  - Sparingly vanishing trigonometric polynomial generating function $f(x) = 2 - 2 \cos x$, ill-conditioned, Zero-valued (order 2) at the origin [42].
- Test 4 $(a_0, a_1, \ldots , a_{n-1})^T = (20, -15, 6, -1, 0, \ldots , 0)^T$.
  - Sparing vanishing trigonometric polynomial generating function $f(x) = (2 - 2 \cos x)^2$, ill-conditioned, Zero-valued (order 6) at the origin.
- Test 5 $(a_0, a_1, \ldots , a_{n-1})^T = (\pi^2/2, -2, 0, -2/9, 0, -2/25, 0, \ldots , 0, -2/k^2, 0, \ldots )^T$.
  - Sparing vanishing generating function $f(x) = \pi |x|$, ill-conditioned, Zero-valued (order 1) at the origin [43].
- Test 6 $(a_0, a_1, \ldots , a_{n-1})^T = (2, 0, 2, 0, -2, 1/5, 0, 2, \ldots , 0, (1)^{1+k^2}/2/(k^2 - 1), 0, \ldots )^T$.
  - Sparing vanishing generating function $f(x) = \pi |\cos x|$, ill-conditioned, Zero-valued (order 1) at $\pi$ [43].

We notice that the generating function $f$ is strictly positive in the two (well-conditioned for $g = 1$) test cases 1 and 2, and sparsely vanishing in the four (ill-conditioned) test cases 3, 4, 5 and 6. However, for $g \geq 2$ we also remind the crucial point that the corresponding $g$-Toeplitz structures are always extremely ill-conditioned, no matter which symbol $f$ is chosen, since the large fraction $(g - 1)/g$ of their singular values always vanish to zero according to (7).

On the ground of Remark 5.4, for any $g$-Toeplitz test matrix we consider both the Natural $g$-circulant preconditioner and the Optimal $g$-circulant preconditioner (see [41,14] for the classical Toeplitz case $g = 1$). The numerical test have been developed with Matlab, and the singular value decomposition has been computed by the built-in Matlab function svd().
6.1. The distribution of the singular values

First, we plot the distribution of the singular values of the $n \times n$ $g$-Toeplitz matrix $A$, the corresponding $g$-circulant preconditioner $P$, and the preconditioned matrix $P^{-1}A$, for $n = 1000$ and $g = 2, 3, 7$ ($n$ and $g$ are co-prime for $g = 3$ and $g = 7$, and are not co-prime for $g = 2$). In particular, we have:

(I) Figs. 1 and 2 show the singular values of the $g$-Toeplitz matrices $A$, the Natural (top) and Optimal (bottom) $g$-circulant

---

Fig. 1. $g = 3$ (coprime case)—Singular values of $g$-Toeplitz matrices $A$, Natural (top) and Optimal (bottom) $g$-circulant preconditioners $P$ and corresponding preconditioned matrices $P^{-1}A$. 

---

Natural $g$-circulant preconditioning

Optimal $g$-circulant preconditioning
preconditioners $P$ and the corresponding preconditioned matrices $P^{-1}A$ in the coprime cases, respectively for $g = 3$ and $g = 7$;

(ii) Fig. 3 shows the singular values of the optimal preconditioning in the non-coprime case $g = 2$, for two test examples (Test 1 and Test 5).
Before dealing with the preconditioned matrix $P^{-1}A$, it is quite interesting to notice that the plotted distribution of the singular values of the $g$-Toeplitz matrix $A$ and its $g$-circulant preconditioner $P$ “exactly” agrees with the corresponding expected distributions (6)–(11). Indeed, for $g > 1$ and sparsely vanishing generating functions, we have:

(i) regarding the $g$-Toeplitz matrix $A$, the first $\left\lceil \frac{n}{g} \right\rceil$ singular values are positive, and equal to $\sqrt{\|f\|^{(2)}(x)}$, while the remaining $n - \left\lceil \frac{n}{g} \right\rceil$ vanish to zero, as stated by in Section 4.1 (see the dashed line in Figs. 1–3);

(ii) regarding the $g$-circulant preconditioner $P$, if $\gamma = \gcd(n, g) = 1$ then the singular values are bounded away from zero or sparsely vanishing as well as the generating function is (see the dotted line in Figs. 1 and 2), and, if $\gamma > 1$, the first $\frac{n}{\gamma}$ singular values are always bounded away from zero (regardless of the sparsely vanishing generating function is or is not bounded away from zero) and equal to $\sqrt{\|h\|^{(3)}(x)}$, while the remaining $n - \frac{n}{\gamma}$ are null, as stated Section 4.2 (see the dotted line in Fig. 3).

In particular, since $n = 1000$, then $\gamma = 1$ for $g = 3, 7$, and $\gamma = 2$ for $g = 2$: in Figs. 1 and 2, related to $\gamma = 1$, the singular values (dotted lines) of both the natural and optimal $g$-circulant preconditioners are bounded away from zero in the well-conditioned test cases 1 and 2, and sparsely vanishing in the ill-conditioned test cases 3, 4, 5, and 6, while one half of their singular values are always null in Fig. 3, related to $\gamma = 2$.

It is now interesting to analyze the distribution of the preconditioned matrix.

Any coprime case (Figs. 1 and 2, solid line) gives rise to a good clustering at unity, in the first $\left\lfloor \frac{n}{g} \right\rfloor$ singular values, while the remaining ones are null. This is a result which was expected in the light of Theorem 5.1: the preconditioned matrix $P^{-1}A$ guarantees a good clustering in a subspace which is the most large possible (remember that the rank of $A$ “goes” to $\left\lfloor \frac{n}{g} \right\rfloor$, since $n - \left\lfloor \frac{n}{g} \right\rfloor$ singular values vanish as $n$ increases, so that the rank of $P^{-1}A$ cannot be larger than $\left\lfloor \frac{n}{g} \right\rfloor$). This good clustering at unity of the preconditioned matrix $P^{-1}A$ occurs in both the well-conditioned case (see, in Figs. 1 and 2, the cases Test 1 and Test 2, where the preconditioners have no vanishing singular values) and the ill-conditioned case (see, again in Figs. 1 and 2, the cases from Test 3 to Test 6, where the preconditioners have always a zero measure vanishing singular subspace). We can also observe that the singular value distributions of the natural preconditioned matrix and the optimal preconditioned matrix are very similar. This agrees with the classical and widely studied Toeplitz case (i.e., $g = 1$), where both the distributions tend to the generating function, as $n$ grows. However, as expected, the optimal preconditioner, which is the closest-to-$A$ $g$-circulant matrix w.r.t. the Frobenius distance, gives a bit better clustering than the natural one: as an instance, see in particular the clustering at unity of Test 3 in the optimal preconditioning (bottom) and in the natural preconditioning (top) in Figs. 1 and 2.

The situation is different in the non-coprime case, as Fig. 3 shows. Before going on, according to Section 4.2, we mention that in this case instead of the inverse $P^{-1}$ we have to consider the Moore–Penrose generalized inverse $P^\dagger$, $P$ being a singular
matrix. Due to the non-coprime circularity, now the $g$-circulant preconditioner has a lot of cyclically repeated, hence linearly dependent, columns. Heuristically, the $g$-circulant preconditioner $P$ "looses" a lot of information which was contained in the related $g$-Toeplitz matrix $A$, which means that $P$ becomes less correlated to $A$, and a good clustering is no longer possible. This is well explained by the first two columns of Fig. 3, solid line, where just a couple of test examples related to the optimal preconditioner are reported (all the others behave similarly, so they are not reported). In particular, in the first two columns we can see that the singular values of the preconditioned matrix $P^TA$ are not clustered (moreover they tend to diverge, giving rise to high instability in real applications). To avoid such an amplification, instead of using $P^T$ for the preconditioned matrix, we can consider a regularized version $P^*_1$ of $P^T$, where the singular values of $P$ smaller than a regularization parameter $\alpha > 0$ are not inverted. As a first attempt, in the third column of Fig. 3, we plot the singular values of the preconditioned matrix $P^TA$, with $\alpha = 10^{-12}\|P\|_2$. As we can notice, a useful clustering is found also for the non-coprime case. However, in the following subsection we study how a regularized version of $g$-circulant preconditioners is useful also in the coprime case, to allow the preconditioned conjugate gradient method to speed up the convergence without suffering from data noise.

6.2. The preconditioning effectiveness

In this subsection we give a first evaluation of the behavior of the optimal $g$-circulant preconditioning for the solution of the $g$-Toeplitz linear system $Ax = b$, with $g = 3 > 1$. First of all we mention that, since the square $g$-Toeplitz matrix $A$ has a large vanishing subspace, we necessarily compute the least square solution $A^*Ax = A^*b$. Accordingly, we consider the solution of the (preconditioned) conjugate gradient method applied to the normal equations.

In order to evaluate the restoration errors, we choose the true data vector $x$, ad then we compute the known data $b$ simply as $b = Ax$. Let $x_k$ be the $k$-th iteration of the (preconditioned) conjugate gradient method on the normal equations. We compute the relative residual error $\text{RelRes} = \frac{\|A^*Ax_k - A^*b\|}{\|A^*b\|}$ and the relative error on the restored signal $\text{RelErr} = \frac{\|x_k - x^t\|}{\|x^t\|}$, where $x^t$ is the projection of the true data on the complementary of the vanishing subspace (which is obviously the best possible restoration). Since the rank of $A$ tends to $\lfloor n/g \rfloor$ as $n$ increases, to obtain $x^t$ we compute $x^t = \tilde{V}\tilde{V}^*x$, where $\tilde{V}$ is the $n \times \lfloor n/g \rfloor$ matrix given by the first $\lfloor n/g \rfloor$ columns of $V$, $V$ being the orthogonal matrix of the singular value decomposition $A = U\Sigma V^*$.

By using the built-in Matlab function pcg() within the maximum number of iterations equal to 100 and tolerance equal to $10^{-20}$, in Table 1 the numerical results related to three different levels of noise on the data $b$ are reported. In particular, by denoting as $b_\eta = b + \eta$ the noisy data, where $\eta$ is a white Gaussian noise, we have the following test cases: in the left columns the data $b$ is noiseless; in the central columns the relative noise level $\|b_\eta - b\|/\|b\|$ is $10^{-4}\%$; in the central columns the relative noise level $\|b_\eta - b\|/\|b\|$ is $10^{-1}\%$. Here the true data vector $x$ is given by $\cos(g\pi i/n)$, so that the first $n/g$ values of the true data are a sampling on a uniform grid of an entire semi-period of the cosine function.

The result of Table 1 can be summarized as follows: for $g > 1$ the optimal $g$-circulant preconditioners never works (also in the noiseless case), since the optimal $g$-circulant preconditioned conjugate gradient method never allows us to obtain better results than the "unpreconditioned" method. In the classical Toeplitz case, this is a well known fact for ill-conditioned cases (here the last four test cases 3–6). Indeed, most preconditioners for Toeplitz systems with high clustering of the singular values such as the natural and optimal ones give rise to instability and noise amplification [39], and now for $g > 1$ we can say that this phenomenon is even amplified. However, the new important fact is that $g$-Toeplitz matrices are always ill-conditioned for $g > 1$, also when the classical Toeplitz matrix (i.e., $g = 1$) is well-conditioned. This means that, although for
Table 1

<table>
<thead>
<tr>
<th>Noise level</th>
<th>Preconditioning</th>
<th>0 (No noise)</th>
<th>0.0001%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iter. number</td>
<td>92</td>
<td>48</td>
<td>94</td>
<td>77</td>
</tr>
<tr>
<td>Relative residual</td>
<td>3.19e−007</td>
<td>6.05e−010</td>
<td>3.16e−007</td>
<td>1.64e−009</td>
</tr>
<tr>
<td>Relative error</td>
<td>4.29e−006</td>
<td>1.20e−008</td>
<td>1.89e−005</td>
<td>1.88e−005</td>
</tr>
<tr>
<td>Test 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iter. number</td>
<td>47</td>
<td>13</td>
<td>47</td>
<td>36</td>
</tr>
<tr>
<td>Relative residual</td>
<td>2.87e−010</td>
<td>2.89e−010</td>
<td>3.48e−010</td>
<td>8.94e−010</td>
</tr>
<tr>
<td>Relative error</td>
<td>4.64e−010</td>
<td>4.72e−010</td>
<td>8.75e−006</td>
<td>8.75e−006</td>
</tr>
<tr>
<td>Test 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iter. number</td>
<td>100⁺</td>
<td>2</td>
<td>100⁺</td>
<td>2</td>
</tr>
<tr>
<td>Relative residual</td>
<td>0.0214</td>
<td>1.10e−016</td>
<td>0.0213</td>
<td>1.12e−016</td>
</tr>
<tr>
<td>Relative error</td>
<td>0.0181</td>
<td>1.19e−016</td>
<td>0.0181</td>
<td>2.98e−006</td>
</tr>
<tr>
<td>Test 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iter. number</td>
<td>100⁺</td>
<td>14</td>
<td>100⁺</td>
<td>20</td>
</tr>
<tr>
<td>Relative residual</td>
<td>1.93e−005</td>
<td>3.78e−010</td>
<td>1.93e−005</td>
<td>8.39e−010</td>
</tr>
<tr>
<td>Relative error</td>
<td>7.48e−006</td>
<td>2.40e−010</td>
<td>7.62e−006</td>
<td>2.20e−006</td>
</tr>
<tr>
<td>Test 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iter. number</td>
<td>100⁺</td>
<td>8</td>
<td>100⁺</td>
<td>37</td>
</tr>
<tr>
<td>Relative residual</td>
<td>6.08e−005</td>
<td>1.04e−010</td>
<td>5.97e−005</td>
<td>9.93e−001</td>
</tr>
<tr>
<td>Relative error</td>
<td>3.22e−005</td>
<td>8.93e−011</td>
<td>3.17e−005</td>
<td>2.76e−006</td>
</tr>
<tr>
<td>Test 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iter. number</td>
<td>77</td>
<td>4</td>
<td>75</td>
<td>10</td>
</tr>
<tr>
<td>Relative residual</td>
<td>1.92e−004</td>
<td>2.64e−013</td>
<td>1.93e−004</td>
<td>6.14e−010</td>
</tr>
<tr>
<td>Relative error</td>
<td>1.04e−004</td>
<td>2.65e−013</td>
<td>1.05e−004</td>
<td>7.57e−006</td>
</tr>
</tbody>
</table>

$g = 3$: best relative residual $\| A^*b_0 - A^*b_1 \|$, with corresponding iteration number $k$ and relative restoration error $\| x_0 - x_k \| / \| b \|$, with respect to different noise levels $\delta = \| b - b_0 \| / \| b \|$, where $b_0$ is the pre-conditioned Toeplitz case, according to the remarks of Section 5.2.

$g = 1$ in the well-conditioned case the optimal circulant preconditioner is a good tool for Toeplitz systems, this is no longer true for $g > 1$. Basically, the large vanishing subspace of the g-Toeplitz matrices leads always to severely ill-conditioning.

Another interesting remark is that the vanishing subspace of any g-Toeplitz matrix is localized in space (and not only in frequency, which is the case related to classical Toeplitz matrices), since the last $n - \lceil n/g \rceil$ components of $Ax$ vanishes as $n$ increases. This is confirmed by Fig. 4 top, where some singular vectors $v_i$ of the g-Toeplitz matrix are plotted for $n = 100$ and $g = 3$. For $i \leq \lceil n/g \rceil$ all the last $n - \lceil n/g \rceil$ components of the singular vectors $v_i$ vanish (see the first two figures on the top), while the situation is the opposite for $i > \lceil n/g \rceil$ (see the last two figures on the top). For the corresponding g-circulant optimal preconditioner, such a localization does not arise, as shown by the bottom of Fig. 4, where the singular vector $v_i$ of the g-circulant optimal preconditioner are plotted.

However, although g-circulant preconditioners cannot be used, they are a valid basic scheme which can be easily regularized by spectral filtering. As well as for the classical Toeplitz case with $g = 1$, to improve the results (that is, speed up the convergence, without amplifying the instability due to noise or floating point computation), a wide range of regularization techniques can be applied to g-circulant preconditioning. In this direction, the g-circulant preconditioner can be considered as a basic tool for introducing regularization features, which could provide both speed-up and stability to the pcg() method. To show this good tool, we use a basic 1D real model for signal and image deblurring, where the Toeplitz matrix is a Gaussian point spread function with zero mean and variance equal to 5 has been computed by the psfGauss() Matlab function.

In Fig. 5, for $n = 1000$ and $g = 3$, we show some restorations (iterations 2, 5, 10 and 20, from top to bottom) of both the conjugate gradient and the preconditioned conjugate gradient methods, where the optimal g-circulant preconditioners are regularized by means of a Tikhonov filtering. The dashed line is the restoration of the preconditioned conjugate gradient method, the solid line is the restoration of the preconditioned one, and the dotted line is the true solution, which is the projection on the positive values of the sine function with period $[2\pi/7g]$. We plot only the first $\lceil n/g \rceil$ components of the remaining part lives in the vanishing subspace, as already remarked. Any singular value $\sigma_i$ of the g-circulant preconditioner $P$ is shifted by a regularization parameter $\alpha$, that is, the corresponding singular value of the g-circulant regularized preconditioner $P_\alpha$ is $\sigma_i + \alpha$. In 5 we have three different levels of noise, 0%, 1% and 5% (from left to right), and the corresponding regularization parameters are $\alpha = 10^{-6}$, $\alpha = 10^{-4}$ and again $\alpha = 10^{-4}$. The spectral condition numbers are $\mu_2(P) = 4 \cdot 10^{15}$, $\mu_2(P_{10^{-6}}) = 1 \cdot 10^6$ and $\mu_2(P_{10^{-4}}) = 1 \cdot 10^4$. In this respect, the regularization parameter $\alpha$ is a key step in order to obtain a regularizing preconditioner for g-Toeplitz systems, as already noticed for the classical ill-conditioned Toeplitz case, according to the remarks of Section 5.2. As we can notice, now the g-circulant preconditioned
conjugate gradient method outperforms the unpreconditioned one, since the convergence speeds up without high noise amplification. The convergence is really good since just the second iteration of the pcg() method (first row) is good. This is also well explained in the following Fig. 6, where the convergence histories of the relative residual error (top) and the relative error (bottom) has been plotted for the first 20 iterations. As we can see, the relative residual error always decrease with high stability, while the relative error shows the expected semi-convergence behavior when the noise is added. The semi-convergence is well amplified in the 5% noise case: in the last row of Fig. 5 the 20-th iteration of the pcg() gives rise to noise amplifications, and many oscillations occur, and an early stop of the iterations prevents this amplification.

7. Conclusions

In this paper we have studied in detail the singular values of matrix sequences obtained by preconditioning g-Toeplitz sequences associated with a given integrable symbol via g-circulant sequences. The generalization to the multilevel block
Fig. 6. Relative residuals (top) and relative errors (bottom), with three different levels of noise 0%, 1% and 5% (from left to right), with $n = 1000$ and $g = 3$ (cf. Fig. 5).  

setting has been sketched. The main point is that the standard preconditioning works only in the classical setting, namely when $g_i = \pm 1$, $i = 1, \ldots, d$. However, when $g$ (or $|g|$) is positive a basic pre conditioner for regularizing techniques can be obtained by a clever choice of the $g$-circulant sequence $\{C_n^g\}$. We have presented and discussed various numerical results, also instructive for specific applications in image deblurring and denoising. In particular they have confirmed that the proposed preconditioners can be used as a basic tool for obtaining regularizing features, by means of filtering techniques which will be better analyzed and discussed in future works. Indeed, while a positive feature is the selective clustering of the preconditioned sequence, a negative aspect which can be understood also from the numerical experiments is that the subspaces associated with degenerating singular values has a non-trivial intersection with the low frequencies, where usually the signals/images live and the latter is independent of the analytical features of the symbol $f$. Consequently, the problems $\{A_n x = b\}$ are inherently highly ill-posed and this cannot be substantially changed by any choice of the preconditioning sequence. However, spectral filtering can be used on the preconditioned sequences, allowing for effective regularizing strategy.

7.1. Two-dimensional $g$-Toeplitz matrices for structured shift-variant image deblurring

We conclude the paper by briefly introducing a real problem of image deblurring [44] which is related to $g$-Toeplitz matrices. Basically, a blurring model (i.e., the forward model) involves a Fredholm linear operator of the first kind as follows. A blurred version $b \in L^2(\mathbb{R}^2)$ of a true image $x \in L^2(\mathbb{R}^2)$ is given by

$$\begin{align*}
    b(v) &= \int_{\mathbb{R}^2} h(v, u)x(u) \, du \quad (20)
\end{align*}$$

where the integral kernel $h \in L^2(\mathbb{R}^2 \times \mathbb{R}^2)$ is the known impulse response of the blurring system, also called a point spread function (PSF), $v = (v_1, v_2)$ and $u = (u_1, u_2)$ being the system coordinates of the blurred image $b$ and the true image $x$. Image deblurring is the (inverse) problem of finding (an approximation of) the true data $x$ (i.e., the cause) by means of the knowledge of the blurred data $b$ (i.e., the effect).

Among all the shift-variant imaging systems, we consider the ones which are intrinsically shift-invariant as follows: there exist two "coordinate transformations" $\phi = \phi(v) \text{ and } \psi = \psi(u)$ such that

$$\begin{align*}
    h(v, u) &= \tilde{h}(\phi(v), \psi(u)),
\end{align*}$$

where $\tilde{h}(\phi, \psi) = h(\phi - \psi)$ is a shift-invariant PSF. In the simplest case where the coordinate transformations are linear functions such as $\phi(v) = v$ and $\psi(u) = gu$, with $g$ integer value, by using a fixed discretization step $d$ we have that

$$\begin{align*}
    A_{ij} &= h(id, jd) = h(i\phi(id) - j\psi(jd)) = h(id - jgd). \quad (21)
\end{align*}$$
This means that the PSF matrix $A$ is a $g$-Toeplitz matrix. However, in general, we have to consider $(g, f)$-Toeplitz matrices, that is, matrices which obey the rule $A_n = \left[ a_{g-f} \right]_{s=0}^{n-1}$, which are direct generations of $g$-Toeplitz matrices. By recalling that any 3D geometric projectivity is a linear transformation, we have that such $(g, f)$-Toeplitz matrices arise in many imaging systems related to large scenes, where the projective geometry becomes important due to the perspective. As an instance $(g, f)$-Toeplitz blurring matrices arise when some objects are moving with approximately the same speed in a plane which is not parallel to the image plane of the imaging apparatus (this is usually called a “non-perpendicular imaging system geometry”). We remark that this is the classical scenario of highway traffic flow control systems.

A very first numerical simulation is shown in Fig. 7, where a structured shift-variant blurred image related to a synthetic homography (i.e., a projectivity between two planes) has been used (see the shift-variant blur which corrupts the image on the left). Since a homography is a linear transformation w.r.t. the homogeny coordinates, the discretization gives rise to two-level $(g, v)$-Toeplitz matrices. By using the involved algebraic structure, the deblurring process can be done within $O(n^2 \log n)$ and in the classical convolutive (i.e. Toeplitz) case. In Fig. 7, center, we show the projectivity under which the blur becomes shift-invariant, which is modeled by a linear transformation of coordinates (see that the same blur is over all the domain of the image on the center, that is, in this image the blur has been transformed into shift-invariant). By means of such a shift-invariant blurred projected image, we can obtain the deblurred image (left image), by using $O(n^2 \log n)$ computation. For this application, the theory of $g$-circulant preconditioning for $g$-Toeplitz systems will be the basis for speeding up the whole restoration process.

Acknowledgments

The authors thank the anonymous referee for detailed and valuable comments which helped to greatly improve the quality of this paper.

The work of CE, SS-C and DS is supported by PRIN 2008 N. 20083KLJEZ.

References