How to prove that a preconditioner can not be optimal

S. Serra Capizzano

Dipartimento di Energetica,
Via Lombroso 6/17, 50134 Firenze, Italy;
Dipartimento di Informatica,
Corso Italia 40, 56100 Pisa, Italy)
(serra@mail.dm.unipi.it)

and

E. Tyrtyshnikov [1]

Institute of Numerical Mathematics,
Russian Academy of Sciences,
Gubkina 8, Moscow 117333, Russia
(tee@inm.ras.ru)

ABSTRACT

In the general case of multilevel Toeplitz matrices, we proved recently that any multilevel circulant preconditioner is not optimal (a cluster it may provide cannot be proper). The proof was based on the concept of quasi-equimodular matrices. However, this concept does not apply, for example, to the sine-transform matrices. In this paper, with a new concept of *partially equimodular* matrices, we cover now all trigonometric matrix algebras widely used in the literature. We propose a technique for proving the non-optimality of certain frequently used preconditioners for some representative sample multilevel matrices. At the same time, we show that these preconditioners are the best, in a certain sense, among the suboptimal preconditioners (with only a general cluster) for multilevel Toeplitz matrices.

---

[1] The work of this author was supported by the Russian Fund of Basic Research (under grant 97-01-00155) and Volkswagen-Stiftung.

1 Introduction

Although a preconditioner can be generally regarded as an approximant to a given matrix, the approximation here is understood in a fairly broad sense. In particular, given C_n and A_n, both of order n, assume that for any $\varepsilon > 0$ there exist matrices $E_{n\varepsilon}$ and $R_{n\varepsilon}$ such that

$$A_n - C_n = E_{n\varepsilon} + R_{n\varepsilon}, \quad ||E_{n\varepsilon}||_2 \leq \varepsilon, \quad \operatorname{rank} R_{n\varepsilon} \leq r(n,\varepsilon) = o(n). \qquad (*)$$

Then let us say that C_n and A_n are ε-close by rank with the rank bound $r(n,\varepsilon)$. If $\gamma_n(\varepsilon)$ counts how many singular values $\sigma_{in}(A_n - C_n)$ are greater than ε, then $(*)$ amounts to the claim that $\gamma_n(\varepsilon) = o(n)$; in other words, the singular values of $A_n - C_n$ have a general cluster at zero. By definition, it becomes a proper cluster if $\gamma_n(\varepsilon) = O(1)$, which holds equally with $r(n,\varepsilon) = r(\varepsilon) = O(1)$, for any $\varepsilon > 0$.

To estimate the convergence rate of the cg-like methods, one should be interested in the ε-closeness of $C_n^{-1}A_n$ and I_n rather than C_n and A_n. In the former case, C_n is called *optimal* for A_n if $r(n,\varepsilon) = O(1)$, and *suboptimal* if $r(n,\varepsilon) = o(n)$. For the cg-like methods, optimal preconditioners provide the superlinear convergence (see [1, 15, 19]).

However, it is easier and still useful to get on with the ε-closeness of C_n and A_n. In this case, C_n is optimal for A_n in the above sense so long as $r(n, \frac{\varepsilon}{||C_n^{-1}||_2}) = O(1)$. More precisely, the following simple but useful proposition holds.

Proposition 1.1 *If C_n and C_n^{-1} are bounded uniformly in n, then A_n and C_n are ε-close by $O(1)$ rank iff $C_n^{-1}A_n$ and I_n are.*

Proof. It is enough to observe that $A_n - C_n = C_n(C_n^{-1}A_n - I_n)$ and $C_n^{-1}A_n - I_n = C_n^{-1}(A_n - C_n)$. Therefore $A_n - C_n$ can be written as the sum of a term of norm bounded by ε and a term of constant rank iff the same thing can be done for $C_n^{-1}A_n - I_n$ for n large enough. □

Notice that if C_n is uniformly bounded in n but we do not suppose anything about C_n^{-1}, then the fact that the singular values of $C_n^{-1}A_n$ are properly clustered at one implies that A_n and C_n are ε-close by $O(1)$ rank. Conversely, if C_n is bounded and A_n and C_n are not ε-close by $O(1)$ rank, then the singular values of $C_n^{-1}A_n$ cannot be properly clustered at one and, therefore, C_n cannot be an optimal preconditioner for A_n. These remarks are now resumed in the next proposition.

Proposition 1.2 *Let C_n be nonsingular. If C_n is bounded uniformly in n and A_n and C_n are not ε-close by $O(1)$ rank, then C_n is not optimal for A_n.*

Proof. By contradiction, if C_n is an optimal preconditioner for A_n, then $C_n^{-1}A_n - I_n$ can be written as a term of norm bounded by $\frac{\varepsilon}{\|C_n\|_2}$ and a term of rank bounded by a constant independent of n. Therefore, $A_n - C_n = C_n(C_n^{-1}A_n - I_n)$ is the sum of a term of norm bounded by ε and a term of constant rank, and this contradicts the assumption that A_n and C_n are not ε close by $O(1)$ rank. □

Proposition 1.1 gives a criterion to establish the existence of an optimal preconditioner by analyzing the difference $A_n - C_n$. However, we have to suppose the uniform of boundedness of C_n and C_n^{-1} and this request is not practical since C_n is unknown. In Proposition 1.2 we give a simpler condition to establish the nonoptimality of a preconditioner C_n but again we suppose something on C_n. In the next proposition we eliminate any assumption on C_n so that the related statement is truly useful to decide the nonoptimality of a preconditioner C_n.

Proposition 1.3 *Let A_n and C_n be nonsingular. If A_n is bounded uniformly in n and if A_n and C_n are not ε-close by $O(1)$ rank then C_n is not optimal for A_n.*

Proof. By contradiction, if C_n is an optimal preconditioner for A_n, then $C_n^{-1}A_n - I_n$ can be written as term of norm bounded by ε and a term of rank bounded by a constant independent of n. Therefore, by using the Sherman-Morrison-Woodbury formula we have that $A_n^{-1}C_n - I_n$ can be written as term of norm bounded by $O(\varepsilon)$ and a term of rank bounded by a constant independent of n. Therefore $-(A_n - C_n) = A_n(A_n^{-1}C_n - I_n)$ is the sum of a term of norm bounded by $O(\varepsilon)$ and a term of constant rank and this contradicts the assumption that A_n and C_n are not ε close by $O(1)$ rank. □

In this paper we propose some technique for proving that

$$r(n,\varepsilon) \neq o(\rho(n)),$$

for certain functions $\rho(n)$, so long as A_n and C_n possess some special properties. In fact, we generalize our previous results from [14]. In that paper we considered C_n of the form $C_n = U_n D_n V_n$, where matrices U_n are unitary, D_n are diagonal, and V_n are unitary quasi-equimodular. For brevity, we refer to C_n as QE-based matrices.

We term a matrix *equimodular* if all its entries are equal in modulus. It is clear that any unitary equimodular matrix of order n has all its entries equal to $1/\sqrt{n}$ in modulus. If there exist two positive constants $0 < c_1 \leq c_2 < \infty$ independent of n so that the entries of a sequence of matrices V_n belong to $[c_1/\sqrt{n}, c_2/\sqrt{n}]$, then the sequence is called *quasi-equimodular*.

A principal result proved in [14] was the following.

Theorem 1.1 *Assume that A_n and C_n are ε-close by rank with the rank bound $r(n,\varepsilon)$ and, for any $\varepsilon > 0$, let A_n have $\rho(n,\varepsilon)$ columns of the Euclidean length not greater than ε. Let $||A_n||_2$ be bounded uniformly in n. Assume also that:*

(a) *the singular values of A_n are not clustered at zero;*

(b) *$C_n = U_n D_n V_n$ are QE-based matrices.*

Then $r(n,\varepsilon) \neq o(\rho(n,\varepsilon))$.

With this theorem we come up with some interesting negative results about optimal preconditioners for multilevel matrices.

Recall that A is a p-level matrix of multiorder $n = (n_1, \ldots, n_p)$ (and of order $N(n) \equiv n_1 \ldots n_p$), if it encompasses $n_1 \times n_1$ blocks, each of them has $n_2 \times n_2$ blocks, and so on. The rows and columns of A can be pointed to through multiindices $i = (i_1, \ldots, i_p)$ and $j = (j_1, \ldots, j_p)$ as follows: $A = [a_{ij}]$, $1 \leq i, j \leq n$, where $1 = (1, \ldots, 1)$, and inequalities between multiindices are understood in the entrywise sense. By definition, $o(n) = o(N(n))$ as $n \to \infty$, and the latter means that *every component of n tends to infinity.*

A_n is called a p-level Toeplitz matrix if $A_n = [a_{i-j}]$, and it is a p-level circulant if $a_{i-j} = a_{i-j\,(\mathrm{mod}\, n)}$, where $i(\mathrm{mod}\, n) \equiv (i_1(\mathrm{mod}\, n_1),\, \dots\, , i_p(\mathrm{mod}\, n_p))$. Toeplitz matrices $A_n = A_n(f)$ are said to be generated by a symbol $f \in L_1(\Pi)$, $z \in \Pi \equiv [-\pi, \pi]^p$, if their elements come from the formal Fourier expansion $f(z) \sim \sum_k a_k\, e^{\mathrm{i}(k,z)}$, where $k = (k_1, \dots, k_p)$, $(k, z) = k_1 z_1 + \dots + k_p z_p$.

For matrices $A_n(f)$, the assumption (a) of Theorem 1.1 can be reformulated in the terms of some properties of the symbol f. We may use the following result: if a symbol $f \in L_1(\Pi)$ is a complex-valued function of $z \in \Pi = [-\pi, \pi]^p$, then the singular values of $A_n(f)$ are *distributed as* $|f(z)|$ with $z \in \Pi$, that is, for any F continuous with a bounded support,

$$\frac{1}{N(n)} \sum_{1 \le i \le n} F(\sigma_i(A_n)) \to \frac{1}{(2\pi)^p} \int_{\Pi} F(|f(z)|) dz.$$

This is a Szego-like theorem obtained in [18]. For the unilevel Hermitian Toeplitz case, the classical distribution results can be found in [8]. It is easy to see that the above assumption (a) is fulfilled so long as $f(z)$ does not vanish on a subset of positive Lebesgue measure in Π. It is equivalent to the demand that $f(z)$ is at most *sparsely vabishing* [16, 5].

For example, consider 2-level Toeplitz matrices $A_n(f)$ for $f(x_1, x_2) = \exp\{\mathrm{i} x_1\}$. In line with the above,

$$A_n = \begin{bmatrix} 0 & I & & \\ & \ddots & \ddots & \\ & & 0 & I \\ & & & 0 \end{bmatrix},$$

where 0 and I are of order n_2 as $n = (n_1, n_2)$. Consider any QE-based matrices C_n that are ε-close to A_n by rank. From Theorem 1.1, since A_n has $\rho(n) = n_2$ zero columns, $r(n, \varepsilon) \ne o(\rho(n))$. Consequently, since $\rho(n) = n_2 \to \infty$ as $n \to \infty$, we conclude that $r(n, \varepsilon) \ne O(1)$. Now, choose any real value s and take up matrices $I_n + sA_n$. With QE-based C_n with $U_n = V_n^*$, it is next to obvious to infer that any preconditioners of the form $I_n + C_n$ for $I_n + sA_n$ are not optimal.

One may remark still that matrices $I_n + C_n$ can be not easily invertible in the case of $V_n \neq U_n^*$ (they might be not QE-based), and hence, we scarcely use them as preconditioners. All the same, in the case where $V_n = U_n^*$ the matrices $I_n + C_n$ remain to be QE-based so long as C_n do. For the above A_n, consider the splitting $A_n = H_{n1} + \mathrm{i}H_{n2}$, where H_{n1} and H_{n2} are Hermitian, and approximate them by QE-based (with common V_n) matrices C_{n1} and C_{n2}, respectively. If the ε-rank bounds for $H_{n1} - C_{n1}$ and $H_{n2} - C_{n2}$ are both $O(1)$, then A_n should be ε-close to $C_n = C_{n1} + \mathrm{i}C_{n2}$ with the rank bound $O(1)$. Since C_n is also QE-based, it contradicts the above negative result. Now, choose a real value s and consider the following Hermitian 2-level Toeplitz matrices: $I_n + sH_{n1}$ and $I_n + sH_{n2}$ (to guarantee that $I_n + sH_{n1}$ and $I_n + sH_{n2}$ are both invertible we may choose $s \in (0,1)$). We have proved that *at least one of the two* does not admit an optimal preconditioner among QE-based matrices of Hermitian pattern. Still, we can not say definitely *which of the two*.

In the above respect, our negative result for the Hermitian case does not look very satisfactory. In pursuit of using Theorem 1.1 in a direct way, we need to produce a symbol giving Hermitian multilevel Toeplitz matrices with sufficiently many zero columns, which is barely possible.

Another criticism of Theorem 1.1 is that it leaves aside some important matrix algebra preconditioners. In particular, the sine-transform matrices [2] of the form

$$S_n = \sqrt{\frac{2}{n}}\ \left[\sin \frac{\pi i j}{n+1}\right]_{ij=1}^{n}$$

are not quasi-equimodular, and hence, prohibited to play as V_n. To cover such cases, we introduce here a new concept of partially equimodular matrices.

Definition. Matrices V_n are called *partially equimodular* if there exist two positive constants c and d independent of n so that, in any column of V_n for any n, the number of entries which are not less in modulus than $c/\sqrt{n}$, is greater than or equal to dn. Matrices $C_n = U_n D_n V_n$, where U_n are unitary, D_n are diagonal, and V_n are unitary partially equimodular will be referred to as PE-based matrices.

In this paper we modify Theorem 1.1 so that matrices C_n are allowed to be PE-based. Since any PE-based matrices are also QE-based, we thus *weaken* Article (b) of the premises. At the same time, we have to *strengthen* Article (a), yet so that it is still equally easy to fulfil for producing the same negative results.

The paper is organized as follows. In Section 2, we discuss the concept of partially equimodular matrices and their relation with trigonometric matrix algebras. We show that all known algebras consist of PE-based matrices. In Section 3, we propose a new technique to study the ε-closeness of matrices. Then, in Section 4, we show how this technique can be applied to the multilevel Toeplitz matrices and in Section 5 we discuss the (more difficult!) Hermitian case.

2 Unitary partially equimodular matrices and matrix algebras

We begin with a very convenient indication for unitary matrices to be partially equimodular.

Theorem 2.1 *Let V_n be unitary $n \times n$ matrices, and assume that the maximal in modulus entry of V_n does not exceed $M/\sqrt{n}$, where M is a constant independent of n. Then matrices V_n are partially equimodular.*

Proof. With $0 < c < M$, consider an arbitrary column of V_n and denote by ζ_n the number of its entries which are not less in modulus than $c/\sqrt{n}$. Since the Euclidean length of the column is equal to 1, we obtain

$$1 \le (n - \zeta_n)\,\frac{c^2}{n} + \zeta_n\,\frac{M^2}{n} \quad \Rightarrow \quad \zeta_n \ge \frac{1 - c^2}{M^2 - c^2}\,n. \ \ \square$$

Consider a sequence of grids $W_n = \{x_{in}\}$ with n nodes $x_{1n} < \ldots < x_{nn}$ defined on a given basic interval I, and suppose that for any n there is a set $\Phi_n = \{\phi_{in}\}$ of n functions orthogonal in the following sense (see [12]):

$$[\phi_{in}, \phi_{jn}]_n \equiv \sum_{l=1}^{n} \left(\phi_{in}\,\bar{\phi}_{jn}\right)(x_{ln}) = \delta_{ij}, \quad 1 \le i, j \le n.$$

It follows that the generalized Vandermonde matrices [7] $V_n = [\phi_{jn}(x_{in})]_{ij=1}^n$ are unitary. By $\{V_n^* D_n V_n\}$, we denote a sequence of matrix algebras, each, for a fixed n, comprizes the matrices of the form $V_n^* D_n V_n$, where V_n is fixed and D_n is an arbitrary diagonal matrix.

When the functions of Φ_n are trigonometric polynomials, we come up with the well-known matrix algebras associated with the classical fast transforms [10, 6, 9]. Canonical examples of such matrix algebras are the circulant, the τ, and the Hartley [3]. The corresponding functions ϕ_{jn} and nodes x_{in} are the following:

$$\begin{array}{ll} \phi_{jn} = \frac{1}{\sqrt{n}} \exp\{\mathrm{i}(j-1)x\}, & x_{in} = \frac{2\pi(i-1)}{n} \in [-\pi, \pi]; \\ \phi_{jn} = \sqrt{\frac{2}{n+1}} \sin(jx), & x_{in} = \frac{\pi i}{n+1} \in [0, \pi]; \\ \phi_{jn} = \frac{1}{\sqrt{n}} \left(\sin((j-1)x) + \cos((j-1)x)\right), & x_{in} = \frac{2\pi(i-1)}{n} \in [-\pi, \pi]. \end{array}$$

In addition, other 7 examples of unitary transforms V_n related to cosine/sine functions are presented in [10]. One more example is the matrix algebra of ε–circulants [4] with $|\varepsilon| = 1$; if $\varepsilon = \exp\{\mathrm{i}2\pi\psi\}$ then the corresponding functions and nodes are of the form

$$\phi_{jn} = \frac{1}{\sqrt{n}} \exp\{\mathrm{i}(j-1+\psi)x\}, \qquad x_{in} = \frac{2\pi(i-1)}{n} \in [-\pi, \pi].$$

In the above cases, it is easy to check that unitary matrices V_n satisfy the hypothesis of Theorem 2.1, and hence, are partially equimodular. The same holds true even in a more general context.

Theorem 2.2 *Assume that the nodes x_{in} are quasi-uniform on I in the sense that $\sum_{i=1}^{n} |\, |I|/n - (x_{i+1\,n} - x_{in})| = o(1)$, and let $\phi_{jn}(x) = \theta_n \phi_j(x)$, where θ_n are constants, and $\phi_j(x)$ are continuous uniformly bounded in j functions with a finite number of zeroes on I. Assume that the matrices $V_n = [\phi_{jn}(x_{in})]$ are unitary. Then, they satisfy the hypothesis of Theorem 2.1.*

Proof. Let $\phi = \phi_1$. It is sufficient to prove that $c_1 n \le \sum_{i=1}^{n} |\phi(x_{in})|^2 \le c_2 n$ for some positive c_1 and c_2. The second inequality follows from the boundedness of ϕ. The first one stems from the demand that the grids are quasi-uniform and ϕ has only a finite number of zeroes. With these assumptions, there is

$\delta > 0$ such that the number of the indices i for which $|\phi(x_{in})| > \delta$ is bounded from below by $c(\delta)n$. Hence, we can take $c_1 = \delta^2 c(\delta)$. Now, since V_n is unitary, we deduce

$$1 = \sum_{i=1}^{n} |(V_n)_{i,1}|^2 = \theta_n^2 \sum_{i=1} |\phi(x_{in})|^2$$

Therefore the relation $1/\sqrt{c_1 n} \leq \theta_n \leq 1/\sqrt{c_2 n}$ is proved and consequently, due to the uniform boundedness of f_j, the inequalities $\max |(V_n)_{i,j}| \leq M/\sqrt{n}$ hold true with M absolute constant. $\square$

Using Theorem 2.1, we now conclude that the matrices V_n in Theorem 2.1 are partially equimodular (PE).

For the p-level case, matrix algebras $\{V_n^* D_n V_n\}$, where $n = (n_1, \ldots, n_p)$, are constructed through the Kronecker products of p sequences of unilevel matrix algebras $\{V_{n_k}^* D_{n_k} V_{n_k}\}$ so that

$$V_n = \bigotimes_{1 \leq k \leq p} V_{n_k}, \quad D_n = \bigotimes_{1 \leq k \leq p} D_{n_k}.$$

Formally, V_{n_k} may correspond to different p sequences of quasi-uniform grids $W_{n_k}^{(k)}$ on I and functions $\Phi_{n_k}^{(k)}$, though we usually assume that there is no dependence on the upper index.

3 Rank bounds for ε-closeness

When choosing C_n to be ε-close to a given A_n, we are interested to know how the rank bounds may behave. For this we propose now a new version of Theorem 1.1 in order to mellow the assumption that C_n are QE-based. We can consider now arbitrary PE-based C_n (rather than QE-based).

Theorem 3.1 *Assume that A_n and C_n are ε-close by rank with the rank bound $r(n, \varepsilon)$ and, for any $\varepsilon > 0$, let A_n have $\rho(n, \varepsilon)$ columns of the Euclidean length not greater than ε. Let $||A_n||_2$ be bounded uniformly in n. Assume also that:*

(a) *the singular values* $\sigma_1(A_n) \geq \ldots \geq \sigma_n(A_n)$ *behave so that for any* $0 < d < 1$, *there is a positive* $q(d)$ *providing that*

$$S(d, A_n) \equiv \sum_{dn \leq j \leq n} \sigma_j^2(A_n) \geq q(d)n$$

for all sufficiently large n*;*

(b) $C_n = U_n D_n V_n$ *are PE-based matrices.*

Then $r(n, \varepsilon) \neq o(\rho(n, \varepsilon))$.

Proof. We may assume that $||C_n||_2$ are bounded uniformly in n. If it is not the case, we pass to another matrices $\tilde{C}_n$ satisfying the same premises with $r(n, \varepsilon)$ being possibly replaced by $2\,r(n, \varepsilon)$. Indeed, the number of singular values for C_n that are greater than $||A_n||_2 + \varepsilon$ can not exceed $r(n, \varepsilon)$, and we may thus get to $\tilde{C}_n$ by cutting off the largest diagonal entries of D_n in the equation $C_n = U_n D_n V_n$.

By contradiction, assume that $r = r(n, \varepsilon) = o(\rho(n, \varepsilon))$ and show, based on this, that for any $\varepsilon > 0$ there must be a column in C_n whose 2-norm is less than or equal to ε for any n large enough.

Let $E_{n\,\varepsilon}$ contain $\rho = \rho(n, \varepsilon)$ the columns of I_n that correspond to ε-small columns of A_n. By contradiction again, let us suppose that for some $\delta > 0$ independent of n, every column of $C_n\, E_{n\,\varepsilon}$ is greater than δ in the 2-norm for infinitely many n. Therefore,

$$\begin{aligned}
\delta\,\rho < ||C_n\, E_{n\,\varepsilon}||_F &= ||A_n + (C_n - A_n)\, E_{n\,\varepsilon}||_F \\
&\leq ||A_n\, E_{n\,\varepsilon}||_F + \|(C_n - A_n)\, E_{n\,\varepsilon}||_F \\
&\leq \varepsilon\rho + \|(C_n - A_n)\, E_{n\,\varepsilon}||_F \\
&\leq \varepsilon\rho + \sigma_1\,(C_n - A_n)r + \varepsilon(\rho - r) \\
&= O\,(r + \varepsilon\,\rho)\,,
\end{aligned}$$

which is at odds with $r = o(\rho)$.

Thus, from $r(n, \varepsilon) = o(\rho(n, \varepsilon))$ we infer that for any $\varepsilon > 0$, for all sufficiently large n there exists a column $e_{\varepsilon\,n}$ of I_n such that $||C_n e_{\varepsilon\,n}||_2 \leq \varepsilon$. Moreover, from $r(n, \varepsilon) = o(n)$, it follows that A_n and C_n have the same clusters [17, 19] and therefore for any positive δ we have

$$S(d, A_n) = S(d, C_n) + o(n).$$

Finally, taking into account that V_n are partially equimodular with the constants c and d, we deduce that

$$\begin{aligned}\varepsilon &\geq ||C_n e_{\varepsilon n}||_2 = ||D_n V_n e_{\varepsilon n}||_2 \\ &\geq \sqrt{\frac{c^2}{n}\, S(d, C_n)} = \sqrt{\frac{c^2}{n}\, (S(d, A_n) + o(n))} \\ &\geq \sqrt{\frac{c^2}{n}\, (q(d)n + o(n)} \geq c\,\sqrt{q(d)} + o(1),\end{aligned}$$

which is impossible due to the arbitrariness of ε. □

Theorem 3.2 *Under the hypotheses of Theorem 3.1, assume that*

$$\lim_{n\to\infty} \rho(n, \varepsilon) = \infty \quad \forall\, \varepsilon > 0.$$

Then the singular values of $A_n - C_n$ can not have a proper cluster at zero.

Proof. If $r(n, \varepsilon)$ is bounded uniformly in n, then $r(n, \varepsilon) = o(\rho(n, \varepsilon))$, which contradicts the conclusion of Theorem 3.1. □

Theorem 3.3 *Under the hypotheses of Theorem 3.2, consider any matrices P_n such that $P_n + A_n$ and $P_n + C_n$ are nonsingular and $P_n + A_n$ is uniformly bounded in n in the spectral norm. Then the matrices $P_n + C_n$ can not be optimal preconditioners for $P_n + A_n$.*

Proof. By direct application of Proposition 1.3, if the singular values of $A_n - C_n$ have not a proper cluster at zero then the preconditioner $P_n + C_n$ is not optimal for $P_n + A_n$. It remains to apply Theorem 3.2. □

In Theorem 3.1, the assumption that A_n has many small columns indicates the direction in which we may seek for negative results. This assumption is not easy to reconcile with some structural requirements on A_n. To this end, it might be useful to present a modification of Theorem 3.1 that allows us to consider A_n with no small columns. A price for this is that we require of V_n a bit more than PE property, though still less than QE property.

Let us say that matrices A_n are *ε-zeroed* on $\rho(n, \varepsilon)$ columns of matrices Z_n if $A_n Z_n$ has $\rho(n, \varepsilon)$ columns of Euclidean length not greater than ε for all sufficiently large n. Also, matrices Z_n will be referred to as *uniformly sparse* if the number of nonzeroes in every column of Z_n is upper bounded uniformly in n.

Theorem 3.4 *Assume that A_n and C_n are ε-close by rank with the rank bound $r(n,\varepsilon)$ and C_n are ε-zeroed on $\rho(n,\varepsilon)$ columns of uniformly sparse unitary matrices Z_n. Let $||A_n||_2$ be bounded uniformly in n. Assume also that:*

(a) *the singular values $\sigma_1(A_n) \geq \ldots \geq \sigma_n(A_n)$ behave so that for any $0 < d < 1$, there is a positive $q(d)$ providing that*

$$S(d, A_n) \equiv \sum_{dn \leq j \leq n} \sigma_j^2(A_n) \geq q(d)n$$

for all sufficiently large n;

(b) *$C_n = U_n D_n V_n$ are PE-based matrices and, in addition, the maximal in modulus entry of V_n does not exceed $M/\sqrt{n}$, where M does not depend on n.*

Then $r(n,\varepsilon) \neq o(\rho(n,\varepsilon))$.

Proof. Denote by k the maximal number of nonzeroes in any column of Z_n. Then, since Z_n are uniformly sparse, $k < +\infty$. Allowing for the entries of V_n being not greater in modulus than $M/\sqrt{n}$, we conclude now that the entries of $Z_n V_n$ in modulus do not exceed $kM/\sqrt{n}$. Since $Z_n V_n$ are unitary as a product of unitary matrices, they are partially equimodular by Theorem 2.1. Now, $\rho(n,\varepsilon)$ columns of $B_n \equiv Z_n^* A_n Z_n$ are ε-small, and it remains to apply Theorem 3.1 to the matrices B_n and arbitrary PE-based C_n. $\square$

3.1 What is the best for multilevel Toeplitz matrices

We start with a remark on the assumption (a) used in Theorem 3.1. In the case of multilevel Toeplitz matrices generated by $f(z)$, it reduces to some assumption on $f(z)$. Note that the inequality $S(d, A_n) \geq q(d)N(n)$, for any $d \in (0,1)$, takes place if $f(z)$ does not vanish in a subset of $[-\pi,\pi]^p$ of positive measure; in other words, $f(z)$ is at most sparsely vanishing [16, 5]. Remarkable that the same property of f accounts for the assumption (a) in the previous Theorem 1.1. It means that the stronger assumption we have introduced does not seem to affect the construction of "bad examples".

Consider a p-variable symbol of the form

$$f(x_1, \ldots, x_p) = \frac{1}{2}\, \exp\{\mathrm{i}k_1 x_1 + \ \ldots\ \mathrm{i}k_p x_p\}, \quad k_j \geq 1,\ j = 1, \ldots, p,$$

and the corresponding p-level Toeplitz matrices $A_n = A_n(f)$, $n = (n_1, \ldots, n_p)$. Since $|f| = 1/2$, the assumption (a) of Theorem 3.1 is fulfilled. Moreover, the number $\rho(n)$ of zero columns of A_n is easily estimated as follows:

$$\rho(n) \ \geq \ c_f\, N(n)\, \sum_{k=1}^{p} \frac{1}{n_k}, \tag{1}$$

where $c_f > 0$ is independent of n. Thus, we come up with the following negative results.

Theorem 3.5 *For $I_n + A_n$, any suboptimal preconditioner of the form $I_n + C_n$, where C_n is a p-level PE -based matrix, provides the singular value cluster for which, for some $c(\varepsilon) > 0$ and infinitely many n, it holds*

$$\gamma_n(\varepsilon) \ \geq \ c(\varepsilon)\, \rho(n), \tag{2}$$

where $\rho(n)$ is defined by (1).

Corollary. *There exist Hermitian p-level Toeplitz matrices for which any suboptimal PE-based preconditioner with $U_n = V_n^*$ provides the singular value cluster with the number of outliers $\gamma_n(\varepsilon)$ subject to (2).*

These results witness that some well-known suboptimal preconditioners are "optimal" for the whole class of multilevel Toeplitz matrices. Let $f(x_1, \ldots, x_p) > 0$ belong to the Wiener class. Then, by a direct extension of R.Chan's arguments for the unilevel case, it was shown in [20] that multilevel circulant preconditioners of G.Strang's and T.Chan's type provide the general clusters with the number of outliers

$$\gamma_n(\varepsilon) \leq k_f\, N(n)\, \sum_{k=1}^{p} \frac{1}{n_k} \tag{3}$$

with $k_f > 0$ independent of n. In [13, 11, 6] it was found that the same yet for T.Chan's circulants holds true for any positive continuous symbol and, what is more, for all known trigonometric matrix algebra preconditioners. Therefore, we can say that this preconditioning technique, unless not very satisfactory for large p, is the best we may count on when PE-based preconditioners $\{V_n^* D_n V_n\}$ are considered.

Note that the trigonometric matrix algebras other than circulants require a different technique [13]. Given a matrix A_n, we choose a preconditioner $C_n = U_n D_n U_n^*$ of Hermitian pattern so that it minimizes $||A_n - C_n||_F$ over all diagonal matrices D_n. It is clear that the minimum is attained at $D_n = \operatorname{diag} U_n^* A_n U_n$. Assume that we have p sequences of grids $W_{n_k} = \{x_{i\, n_k}\}$ and functions $\Phi_{n_k} = \{\phi_{j\, n_k}\}$. Consider the vector functions $\psi_{n_k}(x_k) = [\phi_{1\, n_k}(x_k), \ldots, \phi_{n_k\, n_k}(x_k)]^*$ and, with the notation $x = (x_1, \ldots, x_p)$, set

$$\sigma_n(f; x) = \psi^* A_n \psi, \quad \psi = \bigotimes_{1 \le k \le p} \psi_{n_k}(x_k).$$

Obviously, D_n consists of the values of $\sigma_n(f; x_{i_1\, n_1}, \ldots, x_{i_p\, n_p})$, where $1 \le i_k \le n_k$, $k = 1, \ldots, p$. The crucial observation is now that, in the case of p-level Toeplitz matrices $A_n = A_n(f)$, the ε-closeness of C_n and A_n can be naturally related to the closeness of functions $f(x)$ and $\sigma_n(f; x)$. Since a continuous periodic function is uniformly approximated by trigonometric polynomials, it is sufficient (for continuous symbols) to study the case when f is a trigonometric polynomial.

Instead of the case study of different algebras, when following the above lines it is possible to propose a unifying approach that covers at once all cases of interest. It is based on a matrix interpretation of Korovkin's results in the approximation theory (see [13]) allowing us to study the latter problem only for a finite set of very simple polynomials.

4 Negative results for Hermitian multilevel matrices

Consider PE-based matrices of Hermitian pattern ($U_n = V_n^*$) and symbols of the form

$$f(x_1, \ldots, x_p) = \frac{1}{2} \exp\{\mathrm{i}k_1 x_1 + \ \ldots \ \mathrm{i}k_p x_p\},$$

$k_j \ge 1$, for $j = 1, \ldots, p$. Let $\operatorname{re}(f)$ and $\operatorname{im}(f)$ denote the real and the imaginary part of f. It is obvious that for any real value s at least one of the two Hermitian Toeplitz matrices $A_n(s + \mathrm{re}(f))$ and $A_n(s + \mathrm{im}(f))$ does not admit an optimal PE-based preconditioner. Therefore, for any fixed positive integer k, for any $j = 1, \ldots, p$ and for any real value s one of the two matrices

$$A_n(s + \cos(kx_j)) \quad \text{or} \quad A_n(s + \sin(kx_j))$$

cannot be optimally preconditioned. Still, we can not say definitely *which of the two*. Yet it is reasonable to think that both. If fact, it is easy to see that $A_n(s + \cos(kx_j))$ is similar to $A_n(s + \cos(k(x_j + v))$ through a *unitary diagonal transformation*. Choosing $v = -\frac{\pi}{2k}$ we have $A_n(s+\cos(k(x_j+v)) = A_n(s + \sin(kx_j))$ and, therefore, by supposing that the considered PE-based matrices $V_n^* D_n V_n$ possess no structural symmetries, we do not see a special motivation for which $A_n(s+\sin(kx_j))$ should be better than $A_n(s+\cos(kx_j))$ or vice-versa. It remains to give a rigorous proof of the preceding qualitative argument and this will be the subject of a future research.

References

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

[2] D. Bini and M. Capovani, Spectral and computational properties of band symmetric Toeplitz matrices, *Linear Algebra Appl.*, **52/53** (1983), pp. 99–126.

[3] D. Bini and P. Favati, On a matrix algebra related to the discrete Hartley transform, *SIAM J. Matrix Anal. Appl.*, **14** (1993), pp. 500–507.

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

[5] F. Di Benedetto and S. Serra Capizzano, A unifying approach to abstract matrix algebra preconditioning, *Numer. Math.*, to appear.

[6] F. Di Benedetto and S. Serra Capizzano, Optimal and superoptimal matrix algebra operators, *TR nr. 360, Dept. of Mathematics - Univ. of Genova*, (1997).

[7] W. Gautschi, The condition of Vandermonde-like matrices involving orthogonal polynomials, *Linear Algebra Appl.*, **52/53** (1983), pp. 293–300.

[8] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications.* Second Edition, Chelsea, New York, 1984.

[9] G. Heinig and K. Rost, Representation of Toeplitz-plus-Hankel matrices using trigonometric transformations with applications to fast matrix-vector multiplication, *personal communication*, (1997).

[10] T. Kailath and V. Olshevsky, Displacement structure approach to discrete-trigonometric-transform based preconditioners of G. Strang type and T. Chan type, *Proc. "Workshop on Toeplitz matrices"*, Cortona (Italy), September 1996. *Calcolo*, **33** (1996), pp. 191–208.

[11] S. Serra Capizzano, Toeplitz preconditioners constructed from linear approximation processes, *SIAM J. Matrix Anal. Appl.,* in press.

[12] S. Serra, Korovkin theorems and linear positive Gram matrix algebras approximation of Toeplitz matrices, *Linear Algebra Appl.*, in press.

[13] S. Serra, A Korovkin-type theory for finite Toeplitz operators via matrix algebras, *Numer. Math.*, to appear.

[14] S. Serra Capizzano and E. Tyrtyshnikov, Any circulant-like preconditioner for multilevel Toeplitz matrices is not optimal, *TR nr. 27, LAN - Dept. of Mathematics - Univ. of Calabria*, (1997).

[15] A. van der Sluis and H.A. van der Vorst, The rate of convergence of conjugate gradients, *Numer. Math.* **48** (1986), pp. 543–560.

[16] E. Tyrtyshnikov, Circulant preconditioners with unbounded inverses, *Linear Algebra Appl.*, **216** (1995), pp. 1–23.

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

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

[19] E. Tyrtyshnikov, *A Brief Introduction to Numerical Analysis*, Birkhauser, Boston, 1997.

[20] E. Tyrtyshnikov, Distributions and clusters. In: *Matrix Methods and Algorithms*, Institute of Numerical Mathematics of the Russian Academy of Sciences, 1993, pp. 124–166.