Optimal trigonometric preconditioners for nonsymmetric Toeplitz systems

Daniel Potts and Gabriele Steidl

220/96

Abstract. This paper is concerned with the solution of systems of linear equations $\boldsymbol{T}_N \boldsymbol{x}_N = \boldsymbol{b}_N$, where $\{\boldsymbol{T}_N\}_{N \in \mathbb{N}}$ denotes a sequence of nonsingular nonsymmetric Toeplitz matrices arising from a generating function of the Wiener class. We present a technique for the fast construction of optimal trigonometric preconditioners $\boldsymbol{M}_N = \boldsymbol{M}_N(\boldsymbol{T}_N' \boldsymbol{T}_N)$ of the corresponding normal equation. Moreover, we prove that the spectrum of the preconditioned matrix $\boldsymbol{M}_N^{-1} \boldsymbol{T}_N' \boldsymbol{T}_N$ is clustered at 1 such that the CG-method applied to the normal equation converges superlinearly. Numerical tests confirm the theoretical expectations.

1991 *Mathematics Subject Classification.* 65F10, 65F15, 65T10.
Key words and phrases. Toeplitz matrix, Krylov space methods, CG-method, preconditioners, normal equation, clusters of eigenvalues.

1 Introduction

Consider the system of linear equations

$$\boldsymbol{T}_N \boldsymbol{x}_N = \boldsymbol{b}_N \,, \tag{1.1}$$

where $\boldsymbol{T}_N \in \mathbb{R}^{N,N}$ denotes a nonsingular Toeplitz matrix. Toeplitz systems arise in a variety of applications in mathematics and engineering (see [7] and the references therein). While there exist fast direct Toeplitz solvers for Hermitian positive definite Toeplitz matrices $\boldsymbol{T}_N$, such techniques are not available in the non-Hermitian case. Iterative metods like GMRES and CG often provide a fast solution of (1.1) if they are applied in connection with preconditioning techniques [7]. In particular, these methods profit from the fact that the vector multiplication with the Toeplitz matrix $\boldsymbol{T}_N$ in each iteration step can be computed with $O(N \log N)$ arithmetical operations by using the fast Fourier transform (FFT). Clearly, the multiplication with the preconditioned matrix should have the same arithmetic complexity. Two types of preconditioners are mainly exploited for linear Toeplitz systems, namely optimal (Cesáro) circulant preconditioners $\boldsymbol{M}_N = \boldsymbol{C}_N(\boldsymbol{T}_N)$ [5] and more simple so-called "Strang" circulant preconditioners $\boldsymbol{M}_N = \boldsymbol{S}_N(\boldsymbol{T}_N)$ [6]. One reason for the choice of circulant preconditioners is the fact that circulant matrices can be diagonalized by the Fourier matrix $\boldsymbol{F}_N := \left(e^{-2\pi ijk/N}\right)_{j,k=0}^{N-1}$, where the multiplication of a vector with $\boldsymbol{F}_N$ takes only $O(N \log N)$ arithmetical operations. Moreover, under certain assumptions on the generating function of $\boldsymbol{T}_N$ (see [8], [22]), it can be proved that the singular values of $\boldsymbol{M}_N^{-1}\boldsymbol{T}_N$ are clustered at 1. For non-Hermitian $\boldsymbol{T}_N$, this results in a superlinear convergence of the CG-method applied to the system

$$(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)^*(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)\boldsymbol{x}_N = (\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)^*\boldsymbol{M}_N^{-1}\boldsymbol{b}_N \,. \tag{1.2}$$

To our knowledge, up to now, for non-Hermitian Toeplitz systems and Toeplitz least square problems, only circulant preconditioners with respect to Toeplitz matrices were constructed and used in some kind of normal equation as in (1.2) [8], [11] or as so-called displacement preconditioners [12]. When we finished the paper, we became aware of new results of E.E. Tyrtyschnikov et al. concerning the convergence behaviour of the preconditioned GMRES-method [23] which avoids the transition of (1.1) to the normal equation. However, the preconditioners are again (improved) circulants, which were constructed with respect to $\boldsymbol{T}_N$.

In this paper, we restrict our attention to nonsymmetric *real* Toeplitz matrices $\boldsymbol{T}_N$. Here, it seems to be natural, to replace the circulant matrices by matrices which are diagonalizable by some real trigonometric matrices. Of course, the commonly used trigonometric transforms are closely related to the Fourier transform. Indeed, for symmetric Toeplitz matrices $\boldsymbol{T}_N$ with positive continuous 2π–periodic generating functions, trigonometric preconditioning significantly accerelates the convergence of the CG-method.

In this paper, we suggest the solution of (1.1) by applying the CG-method to the preconditioned normal equation

$$\boldsymbol{M}_N^{-1}\boldsymbol{T}'_N\boldsymbol{T}_N\boldsymbol{x}_N = \boldsymbol{M}_N^{-1}\boldsymbol{T}'_N\boldsymbol{b}_N \,,$$

where in contrast to (1.2), $\boldsymbol{M}_N = \boldsymbol{M}_N(\boldsymbol{T}'_N\boldsymbol{T}_N)$ denotes the optimal preconditioner with respect to $\boldsymbol{T}'_N\boldsymbol{T}_N$. We demonstrate that the construction of such optimal preconditioners can be realized with only $O(N \log N)$ arithmetical operations despite the fact that $\boldsymbol{T}'_N\boldsymbol{T}_N$ is no longer a Toeplitz matrix. We prove that under certain assumptions on $\boldsymbol{T}_N$ the eigenvalues of $\boldsymbol{M}_N^{-1}\boldsymbol{T}'_N\boldsymbol{T}_N$ are clustered at 1. Although our approach works in exactly the same way for different trigonometric transforms, we prefer to investigate the DCT–II preconditioner in detail and add only few facts concerning the other trigonometric preconditioners. We hope that our notation makes the approach for other trigonometric preconditioners immediately clear. Numerical tests were performed for the different trigonometric preconditioners. Note that in all examples, our preconditioning was superior over the method (1.2) with an optimal trigonometric preconditioner $\boldsymbol{M}_N(\boldsymbol{T}_N)$ of $\boldsymbol{T}_N$.

This paper is organized as follows: Section 2 contains the basic matrix notation. In Section 3, we study the relations between trigonometric transforms and Toeplitz matrices. In particular, we introduce a method for the fast vector multiplication with real nonsymmetric Toeplitz matrices based on real trigonometric transforms. In Section 4, we introduce optimal trigonometric preconditioners. Section 5 is concerned with the proof that the eigenvalues of the preconditioned matrix $\boldsymbol{M}_N^{-1}\boldsymbol{T}'_N\boldsymbol{T}_N$ are clustered at 1. In Section 6, we present the fast construction of the optimal preconditioner. Finally, Section 7 confirms the theoretical expectations by numerical tests.

2 Notation

For the sake of clarity, we collect the matrix notation in this preliminary section.
Let $\boldsymbol{a}_N := (a_0, \ldots, a_{N-1})'$, $\boldsymbol{b}_N := (b_0, \ldots, b_{N-1})'$ and let $\boldsymbol{0}_N$ be the vector consisting of N zeros. Here $\boldsymbol{A}'$ is the transpose of $\boldsymbol{A}$. By $\boldsymbol{I}_N$ we denote the (N,N)-identity matrix and by $\boldsymbol{e}_k \in \mathbb{R}^N$ the k-th identity vektor. To describe Toeplitz and Hankel matrices, we use the following notation:

$$\text{toeplitz}(\boldsymbol{a}', \boldsymbol{b}') := \begin{pmatrix} a_0 & a_1 & \ldots & a_{N-1} \\ b_1 & a_0 & \ldots & a_{N-2} \\ \vdots & \vdots & \vdots & \vdots \\ b_{N-2} & b_{N-3} & \ldots & a_1 \\ b_{N-1} & b_{N-2} & \ldots & a_0 \end{pmatrix} \quad (\text{with } \; a_0 = b_0)\,,$$

stoeplitz $\boldsymbol{a}'$: symmetric Toeplitz matrix with first row $\boldsymbol{a}'$,
atoeplitz $\boldsymbol{a}'$: antisymmetric Toeplitz matrix with first row $\boldsymbol{a}'$, where $a_0 = 0$,

$$\text{hankel}(\boldsymbol{a}', \boldsymbol{b}') := \begin{pmatrix} a_0 & \ldots & a_{N-2} & a_{N-1} \\ a_1 & \ldots & a_{N-1} & b_{N-2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N-2} & \ldots & b_2 & b_1 \\ a_{N-1} & \ldots & b_1 & b_0 \end{pmatrix} \quad (\text{with } a_{N-1} = b_{N-1})\,,$$

shankel $\boldsymbol{a}'$: persymmetric Hankel matrix with first row $\boldsymbol{a}'$,
ahankel $\boldsymbol{a}'$: antipersymmetric Hankel matrix with first row $\boldsymbol{a}'$, where $a_{N-1} = 0$.

Further, we introduce the matrices

$$\boldsymbol{Z}'_{N,1} := \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{N,N+1}\,,\ \boldsymbol{Z}'_{N,2} = \begin{pmatrix} 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{N,N+1}\,,$$

and

$$\boldsymbol{R}'_N := \begin{pmatrix} 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{N-1,N+1}\,.$$

Let $\operatorname{diag} \boldsymbol{a}$ be the diagonal matrix with diagonal $\boldsymbol{a}$ and let $\delta(\boldsymbol{A}) := \operatorname{diag}(a_{k,k})_{k=0}^{N-1}$, where $a_{k,k}$ is the (k,k)–th entry of $\boldsymbol{A}$. By

$$\operatorname{tr} \boldsymbol{A} := \sum_{k=0}^{N-1} a_{k,k}$$

we denote the *trace* of $\boldsymbol{A}$. Moreover, we need the following *matrix norms*:
Spectral norm:
$||\boldsymbol{A}||_2 :=$ (maximum of the absolute values of the singular values of $\boldsymbol{A}$)$^{1/2}$,
Frobenius norm:
$||\boldsymbol{A}||_F := (\sum_{j,k=0}^{N-1} a_{jk}^2)^{1/2}$,
1*–norm*:
$||\boldsymbol{A}||_1 := \max\{\sum_{j=0}^{N-1} a_{j,k} : k = 0, \dots, N-1\}$.
If it does not make confusion, we use the same notation for the norm of absolute summable sequences $a = \{a_k\}_{k\in\mathbb{Z}} \in l_1$, i.e.

$$||a||_1 := \sum_{k\in\mathbb{Z}} |a_k|\,.$$

3 Trigonometric transforms and Toeplitz matrices

We introduce four discrete sine transforms (DST) and four discrete cosine transforms (DCT) as classified by Wang [24]:

$$\text{DCT–I} \quad : \quad \boldsymbol{C}^I_{N+1} := \left(\frac{2}{N}\right)^{1/2} \left(\varepsilon_j^N \varepsilon_k^N \cos\frac{jk\pi}{N}\right)_{j,k=0}^{N} \in \mathbb{R}^{N+1,N+1},$$

$$\text{DCT–II} \quad : \quad \boldsymbol{C}^{II}_{N} := \left(\frac{2}{N}\right)^{1/2} \left(\varepsilon_j^N \cos\frac{j(2k+1)\pi}{2N}\right)_{j,k=0}^{N-1} \in \mathbb{R}^{N,N},$$

$$\text{DCT–III} \quad : \quad \boldsymbol{C}^{III}_{N} := \left(\boldsymbol{C}^{II}_{N}\right)' \in \mathbb{R}^{N,N},$$

$$\text{DCT–IV} \quad : \quad \boldsymbol{C}^{IV}_{N} := \left(\frac{2}{N}\right)^{1/2} \left(\cos\frac{(2j+1)(2k+1)\pi}{4N}\right)_{j,k=0}^{N-1} \in \mathbb{R}^{N,N},$$

and

$$\text{DST–I} \quad : \quad \boldsymbol{S}^I_{N-1} := \left(\frac{2}{N}\right)^{1/2} \left(\sin \frac{(j+1)(k+1)\pi}{N}\right)^{N-2}_{j,k=0} \in \mathbb{R}^{N-1,N-1},$$

$$\text{DST–II} \quad : \quad \boldsymbol{S}^{II}_N := \left(\frac{2}{N}\right)^{1/2} \left(\varepsilon^N_{j+1} \sin \frac{(j+1)(2k+1)\pi}{2N}\right)^{N-1}_{j,k=0} \in \mathbb{R}^{N,N},$$

$$\text{DST–III} \quad : \quad \boldsymbol{S}^{III}_N := \left(\boldsymbol{S}^{II}_N\right)' \in \mathbb{R}^{N,N},$$

$$\text{DST–IV} \quad : \quad \boldsymbol{S}^{IV}_N := \left(\frac{2}{N}\right)^{1/2} \left(\cos \frac{(2j+1)(2k+1)\pi}{4N}\right)^{N-1}_{j,k=0} \in \mathbb{R}^{N,N},$$

where $\varepsilon^N_k := 1/\sqrt{2}$ $(k = 0, N)$ and $\varepsilon^N_k := 1$ otherwise. We refer to the corresponding transforms as *trigonometric transforms*. It is well-known that the above matrices are orthogonal and that the vector multiplication with any of these matrices takes only $O(N \log N)$ arithmetical operations. Fortunately, there exist implementations of algorithms for the vector multiplication with the above sine and cosine matrices, for example a C-implementation based on [3] and [19].
Moreover, we use the slightly modified DCT–I and DST–I matrices

$$\tilde{\boldsymbol{C}}^I_{N+1} := \left((\varepsilon^N_k)^2 \cos \frac{jk\pi}{N}\right)^{N}_{j,k=0}, \quad \tilde{\boldsymbol{S}}^I_{N-1} := \left(\sin \frac{jk\pi}{N}\right)^{N-1}_{j,k=1}$$

and the slightly modified DCT–III and DST–III matrices

$$\tilde{\boldsymbol{C}}^{III}_N := \left((\varepsilon^N_k)^2 \cos \frac{(2j+1)k\pi}{2N}\right)^{N-1}_{j,k=0}, \quad \tilde{\boldsymbol{S}}^{III}_N := \left((\varepsilon^N_{k+1})^2 \sin \frac{(2j+1)(k+1)\pi}{2N}\right)^{N-1}_{j,k=0}.$$

It holds that

$$\tilde{\boldsymbol{C}}^I_{N+1} \tilde{\boldsymbol{C}}^I_{N+1} = \frac{N}{2} \boldsymbol{I}_{N+1}. \tag{3.1}$$

Theorem 3.1. There exist the following relations between trigonometric transforms and Toeplitz matrices:
i) DCT–I and DST–I:

$$\begin{aligned}
\boldsymbol{R}'_N \boldsymbol{C}^I_{N+1} \boldsymbol{D} \boldsymbol{C}^I_{N+1} \boldsymbol{R}_N &= \frac{1}{2} \text{stoeplitz}(a_0, \ldots, a_{N-2}) + \frac{1}{2} \text{shankel}(a_2, \ldots, a_{N-2}, 0, 0), \\
\boldsymbol{S}^I_{N-1} \boldsymbol{R}'_N \boldsymbol{D} \boldsymbol{R}_N \boldsymbol{S}^I_{N-1} &= \frac{1}{2} \text{stoeplitz}(a_0, \ldots, a_{N-2}) - \frac{1}{2} \text{shankel}(a_2, \ldots, a_{N-2}, 0, 0), \\
\boldsymbol{R}'_N \boldsymbol{C}^I_{N+1} \tilde{\boldsymbol{D}} \boldsymbol{R}_N \boldsymbol{S}^I_{N-1} &= \frac{1}{2} \text{atoeplitz}(0, a_1, \ldots, a_{N-2}) + \frac{1}{2} \text{ahankel}(a_2, \ldots, a_{N-1}, 0), \\
\boldsymbol{S}^I_{N-1} \boldsymbol{R}'_N \tilde{\boldsymbol{D}} \boldsymbol{C}^I_{N+1} \boldsymbol{R}_N &= -\frac{1}{2} \text{atoeplitz}(0, a_1, \ldots, a_{N-2}) + \frac{1}{2} \text{ahankel}(a_2, \ldots, a_{N-1}, 0)
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{D} &:= \text{diag}(d_0, \ldots, d_N), \ \tilde{\boldsymbol{D}} := \text{diag}(0, \tilde{d}_1, \ldots, \tilde{d}_{N-1}, 0), \\
(d_0, \ldots, d_N)' &:= \tilde{\boldsymbol{C}}^I_{N+1} (a_0, \ldots, a_{N-2}, 0, 0)', \\
(\tilde{d}_1, \ldots, \tilde{d}_{N-1})' &:= \tilde{\boldsymbol{S}}^I_{N-1} (a_1, \ldots, a_{N-1})'.
\end{aligned}$$

ii) DCT–II and DST–II:

$$\begin{aligned}
\left(\boldsymbol{C}_N^{II}\right)' \boldsymbol{Z}'_{N,2}\, \boldsymbol{D}\, \boldsymbol{Z}_{N,2}\boldsymbol{C}_N^{II} &= \frac{1}{2}\,\mathrm{stoeplitz}(a_0,\ldots,a_{N-1}) + \frac{1}{2}\,\mathrm{shankel}(a_1,\ldots,a_{N-1},0)\,,\\
\left(\boldsymbol{S}_N^{II}\right)' \boldsymbol{Z}'_{N,1}\, \boldsymbol{D}\, \boldsymbol{Z}_{N,1}\boldsymbol{S}_N^{II} &= \frac{1}{2}\,\mathrm{stoeplitz}(a_0,\ldots,a_{N-1}) - \frac{1}{2}\,\mathrm{shankel}(a_1,\ldots,a_{N-1},0)\,,\\
\left(\boldsymbol{C}_N^{II}\right)' \boldsymbol{Z}'_{N,2}\, \tilde{\boldsymbol{D}}\, \boldsymbol{Z}_{N,1}\boldsymbol{S}_N^{II} &= \frac{1}{2}\,\mathrm{atoeplitz}(0,a_1,\ldots,a_{N-1}) + \frac{1}{2}\,\mathrm{ahankel}(a_1,\ldots,a_{N-1},0)\,,\\
\left(\boldsymbol{S}_N^{II}\right)' \boldsymbol{Z}'_{N,1}\, \tilde{\boldsymbol{D}}\, \boldsymbol{Z}_{N,2}\boldsymbol{C}_N^{II} &= -\frac{1}{2}\,\mathrm{atoeplitz}(0,a_1,\ldots,a_{N-1}) + \frac{1}{2}\,\mathrm{ahankel}(a_1,\ldots,a_{N-1},0)
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{D} &:= \mathrm{diag}(d_0,\ldots,d_N)\,,\ \tilde{\boldsymbol{D}} := \mathrm{diag}(0,\tilde{d}_1,\ldots,\tilde{d}_{N-1},0)\,,\\
(d_0,\ldots,d_N)' &:= \tilde{\boldsymbol{C}}^I_{N+1}\,(a_0,\ldots,a_{N-1},0)'\,,\\
(\tilde{d}_1,\ldots,\tilde{d}_{N-1})' &:= \tilde{\boldsymbol{S}}^I_{N-1}\,(a_1,\ldots,a_{N-1})'\,.
\end{aligned}$$

iii) DCT–IV and DST–IV:

$$\begin{aligned}
\boldsymbol{C}_N^{IV}\, \boldsymbol{D}\, \boldsymbol{C}_N^{IV} &= \frac{1}{2}\,\mathrm{stoeplitz}(a_0,\ldots,a_{N-1}) + \frac{1}{2}\,\mathrm{shankel}(a_1,\ldots,a_{N-1},0)\,,\\
\boldsymbol{S}_N^{IV}\, \boldsymbol{D}\, \boldsymbol{S}_N^{IV} &= \frac{1}{2}\,\mathrm{stoeplitz}(a_0,\ldots,a_{N-1}) - \frac{1}{2}\,\mathrm{shankel}(a_1,\ldots,a_{N-1},0)\,,\\
\boldsymbol{C}_N^{IV}\, \tilde{\boldsymbol{D}}\, \boldsymbol{S}_N^{IV} &= -\frac{1}{2}\,\mathrm{atoeplitz}(0,a_1,\ldots,a_{N-1}) + \frac{1}{2}\,\mathrm{shankel}(a_1,\ldots,a_{N-1},0)\,,\\
\boldsymbol{S}_N^{IV}\, \tilde{\boldsymbol{D}}\, \boldsymbol{C}_N^{IV} &= \frac{1}{2}\,\mathrm{atoeplitz}(0,a_1,\ldots,a_{N-1}) + \frac{1}{2}\,\mathrm{shankel}(a_1,\ldots,a_{N-1},0)
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{D} &:= \mathrm{diag}(d_0,\ldots,d_{N-1})\,,\ \tilde{\boldsymbol{D}} := \mathrm{diag}(\tilde{d}_0,\ldots,\tilde{d}_{N-1})\\
(d_0,\ldots,d_{N-1})' &:= \tilde{\boldsymbol{C}}^{III}_N\,(a_0,\ldots,a_{N-1})'\,,\\
(\tilde{d}_1,\ldots,\tilde{d}_{N-1})' &:= \tilde{\boldsymbol{S}}^{III}_N\,(a_1,\ldots,a_{N-1},0)'\,.
\end{aligned}$$

Note that for fixed diagonal matrices $\boldsymbol{D}, \tilde{\boldsymbol{D}}$, the above decompositions into a Toeplitz and a Hankel matrix are not unique.

Proof: We restrict the proof to the DCT–II. To simplify the notation, we drop the index N and set $\boldsymbol{C} := \boldsymbol{Z}_{N,2}\boldsymbol{C}_N^{II}$, $\boldsymbol{S} := \boldsymbol{Z}_{N,1}\boldsymbol{S}_N^{II}$ and $\boldsymbol{D} := \mathrm{diag}(d_0,\ldots,d_N)$. Then by

$$\cos\alpha\,\cos\beta = \frac{1}{2}\,\cos(\alpha-\beta) + \frac{1}{2}\,\cos(\alpha+\beta)\,,$$

the (u,v)–entry of the matrix $\boldsymbol{C}'\boldsymbol{D}\boldsymbol{C}$ is

$$(\boldsymbol{C}'\boldsymbol{D}\boldsymbol{C})_{u,v} = \frac{1}{2}\,\frac{2}{N}\sum_{k=0}^{N-1}(\varepsilon_k^N)^2\, d_k \cos\frac{(u-v)k\pi}{N} + \frac{1}{2}\,\frac{2}{N}\sum_{k=0}^{N-1}(\varepsilon_k^N)^2\, d_k \cos\frac{(u+v+1)k\pi}{N}\,,$$

or equivalently, since $-(-1)^{u-v} d_N = -(-1)^{u+v+1} d_N$ for arbitary $d_N \in \mathbb{R}$,

$$(\boldsymbol{C}'\boldsymbol{D}\boldsymbol{C})_{u,v} = \frac{1}{2}\frac{2}{N}\sum_{k=0}^{N}(\varepsilon_k^N)^2\, d_k \cos\frac{(u-v)k\pi}{N} + \frac{1}{2}\frac{2}{N}\sum_{k=0}^{N}(\varepsilon_k^N)^2\, d_k \cos\frac{(u+v+1)k\pi}{N}.$$

Choosing $d_N \in \mathbb{R}$ such that

$$\sum_{k=0}^{N}(\varepsilon_k^N)^2\, d_k(-1)^k = 0\,,$$

we get by symmetry properties of cosine function that

$$\boldsymbol{C}'\boldsymbol{D}\,\boldsymbol{C} = \frac{1}{2}\,\mathrm{stoeplitz}(a_0,\ldots,a_{N-1}) + \frac{1}{2}\,\mathrm{shankel}(a_1,\ldots,a_{N-1},0)\,,$$

where

$$(a_0,\ldots,a_{N-1},0)' = \frac{2}{N}\,\tilde{\boldsymbol{C}}^I_{N+1}\,(d_0,\ldots,d_N)'\,,$$

i.e. by (3.1)

$$(d_0,\ldots,d_N)' = \tilde{\boldsymbol{C}}^I_{N+1}\,(a_0,\ldots,a_{N-1},0)'\,.$$

The other decomposition relations follow in a similar way by application of $\sin\alpha\,\sin\beta = \frac{1}{2}\cos(\alpha-\beta) - \frac{1}{2}\cos(\alpha+\beta)$ and $\sin\alpha\,\cos\beta = \frac{1}{2}\sin(\alpha-\beta) + \frac{1}{2}\sin(\alpha+\beta)$. ∎

Theorem 3.1 provides a new method for the fast multiplication of a real vector with a real nonsymmetric Toeplitz matrix that avoids the complex arithmetic which comes into the play if we exploit the usual FFT–based method for the fast vector – Toeplitz matrix multiplication.

Corollary 3.2. (Fast vector multiplication with nonsymmetric Toeplitz matrices) Let

$$\boldsymbol{T} = \boldsymbol{T}_N := (t_{j-k})_{j,k=0}^{N-1} = \mathrm{toeplitz}((t_0,t_{-1},\ldots,t_{-(N-1)}),(t_0,t_1,\ldots,t_{N-1}))$$

be given and let $\boldsymbol{C} := \boldsymbol{Z}_{N,2}\boldsymbol{C}_N^{II}$, $\boldsymbol{S} := \boldsymbol{Z}_{N,1}\boldsymbol{S}_N^{II}$. Then

$$\boldsymbol{T} = \frac{1}{2}(\boldsymbol{T}+\boldsymbol{T}') + \frac{1}{2}(\boldsymbol{T}-\boldsymbol{T}') = \boldsymbol{C}'\boldsymbol{D}\boldsymbol{C} + \boldsymbol{S}'\boldsymbol{D}\boldsymbol{S} + \boldsymbol{C}'\tilde{\boldsymbol{D}}\boldsymbol{S} - \boldsymbol{S}'\tilde{\boldsymbol{D}}\boldsymbol{C}\,,$$

where

$$\boldsymbol{D} := \mathrm{diag}(d_0,\ldots,d_N)\,,\quad \tilde{\boldsymbol{D}} := \mathrm{diag}(0,\tilde{d}_1,\ldots,\tilde{d}_{N-1},0)\,,$$

$$\begin{aligned}(d_0,\ldots,d_N)' &:= \tilde{\boldsymbol{C}}^I_{N+1}\,(t_0,\frac{t_1+t_{-1}}{2},\ldots,\frac{t_{N-1}+t_{-(N-1)}}{2},0)'\,,\\ (\tilde{d}_1,\ldots,\tilde{d}_{N-1})' &:= \tilde{\boldsymbol{S}}^I_{N-1}\,(\frac{t_{-1}-t_1}{2},\ldots,\frac{t_{-(N-1)}-t_{N-1}}{2})'\,.\end{aligned}$$

The vector multiplication with $\boldsymbol{T}$ requires except of $O(N)$ additions
- one DCT–I and one DST–I to build $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$ in a precomputation step,

- one DCT–II and one DST–II,
- four multiplications of vectors with diagonal matrices,
- one DCT–III and one DST–III of the vectors $\boldsymbol{DCx} + \tilde{\boldsymbol{D}}\boldsymbol{Sx}$ and $\boldsymbol{DSx} - \tilde{\boldsymbol{D}}\boldsymbol{Cx}$, respectively, and takes therefore only $O(N \log N)$ arithmetical operations.

Clearly, by Theorem 3.1, we can formulate similar algorithms for the fast multiplication of vectors with Toeplitz or Hankel matrices with respect to the other trigonometric transforms. Typewriting this paper, we got a ps-file of a paper of G. Heinig and K. Rost [14], which contains results in a similar direction as presented in this section.

4 Optimal trigonometric preconditioners

We are concerned with the solution of the system of linear equations

$$\boldsymbol{T}_N \boldsymbol{x}_N = \boldsymbol{b}_N$$

with a nonsingular nonsymmetric Toeplitz matrix $\boldsymbol{T}_N \in \mathbb{R}^{N,N}$. We intend to solve the normal equation

$$\boldsymbol{T}_N' \boldsymbol{T}_N \boldsymbol{x}_N = \boldsymbol{T}_N' \boldsymbol{b}_N \tag{4.1}$$

by the CG-method. In Section 7, we will see that with a good preconditioner at hand, this can be realized in a fast way. There are several requirements on a preconditioner $\boldsymbol{M}_N$ of (4.1) resulting from the construction and the convergence behaviour of the CG-method as well as from the fact that the vector multiplication with $\boldsymbol{T}_N$ requires only $O(N \log N)$ arithmetical operations. Therefore, we are looking for a preconditioner with the following properties:

(P1) $\boldsymbol{M}_N$ is symmetric and positive definite such that the bilinear form

$$(\boldsymbol{x}_N, \boldsymbol{y}_N)_{M_N} := \boldsymbol{x}_N' \boldsymbol{M}_N \boldsymbol{y}_N$$

arising in the left preconditioned CG-method is symmetric and positive definite, too.

(P2) The spectrum of $\boldsymbol{M}_N^{-1} \boldsymbol{T}_N' \boldsymbol{T}_N$ is clustered at 1.

(P3) The vector multiplication with $\boldsymbol{M}_N$ can be computed with $O(N \log N)$ arithmetical operations.

(P4) The construction of $\boldsymbol{M}_N$ takes only $O(N \log N)$ arithmetical operations.

Having property (P3) in mind, a straightforward idea consists in choosing $\boldsymbol{M}_N$ from an algebra

$$\mathcal{A}_{O_N} := \{\boldsymbol{O}_N' (\operatorname{diag} \boldsymbol{d})\, \boldsymbol{O}_N : \boldsymbol{d} \in \mathbb{R}^N\} \tag{4.2}$$

of matrices which are diagonalizable by some orthogonal matrix $\boldsymbol{O}_N$, where $\boldsymbol{O}_N$ has the additional property that its vector multiplication requires only $O(N \log N)$ arithmetical operations. As orthogonal matrices, we will use the trigonometric matrices of the previous section which are closely related to the Fourier matrix $\boldsymbol{F}_N$, but have the advantage of purely real entries. Moreover, if we choose $\boldsymbol{M}_N \in \mathcal{A}_{O_N}$ as so-called optimal preconditioner of $\boldsymbol{T}_N' \boldsymbol{T}_N$, then we will see that under certain assumptions on $\boldsymbol{T}_N$, the properties (P1), (P2) and (P4) are also fulfilled.

For $\boldsymbol{A}_N \in \mathbb{R}^{N,N}$, the matrix $\boldsymbol{M}_N(\boldsymbol{A}_N)$ is called an *optimal preconditioner* of $\boldsymbol{A}_N$ in $\mathcal{A}_{O_N}$ [5] if

$$||\boldsymbol{M}_N(\boldsymbol{A}_N) - \boldsymbol{A}_N||_F = \min\{||\boldsymbol{B}_N - \boldsymbol{A}_N||_F : \boldsymbol{B}_N \in \mathcal{A}_{O_N}\}\,. \tag{4.3}$$

If $\boldsymbol{O}_N$ is one of the orthogonal matrices which correspond to the DST–I, DST–II, DCT–II, DST–IV or DCT–IV, respectively, then $\boldsymbol{M}_N$ is said to be an *optimal trigonometric preconditioner* of $\boldsymbol{A}_N$. The choice of the Frobenius norm in definition (4.3) results from the fact that the Frobenius norm is induced by an inner product of $\mathbb{R}^{N,N}$

$$\langle \boldsymbol{A}_N, \boldsymbol{B}_N \rangle := \mathrm{tr}(\boldsymbol{A}_N' \boldsymbol{B}_N) = \sum_{j,k=0}^{N-1} a_{j,k} b_{j,k}. \tag{4.4}$$

In particular, it holds that

$$||\boldsymbol{O}_N \boldsymbol{A}_N \boldsymbol{O}_N'||_F^2 = \mathrm{tr}(\boldsymbol{O}_N \boldsymbol{A}_N' \boldsymbol{O}_N' \boldsymbol{O}_N \boldsymbol{A}_N \boldsymbol{O}_N') = \mathrm{tr}(\boldsymbol{A}_N' \boldsymbol{A}_N) = ||\boldsymbol{A}_N||_F^2\,. \tag{4.5}$$

The following lemma describes the optimal preconditioner of $\boldsymbol{T}_N' \boldsymbol{T}_N$ in two different ways.

Lemma 4.1. Let $\boldsymbol{A}_N \in \mathbb{R}^{N,N}$ and let $\mathcal{A}_{O_N}$ be defined by (4.2) with respect to some orthogonal matrix $\boldsymbol{O}_N$. Then the optimal preconditioner of $\boldsymbol{A}_N$ is given by

$$\boldsymbol{M}_N(\boldsymbol{A}_N) = \boldsymbol{O}_N' \, \delta(\boldsymbol{O}_N \boldsymbol{A}_N \boldsymbol{O}_N') \, \boldsymbol{O}_N\,. \tag{4.6}$$

If $\{\boldsymbol{B}_0, \ldots, \boldsymbol{B}_{N-1}\}$ denotes a basis of $\mathcal{A}_{O_N}$, then an alternative description of $\boldsymbol{M}_N(\boldsymbol{A}_N)$ reads as

$$\boldsymbol{M}_N(\boldsymbol{A}_N) := \sum_{k=0}^{N-1} \alpha_k \boldsymbol{B}_k\,, \tag{4.7}$$

where the coefficient vector $\boldsymbol{\alpha} := (\alpha_0, \ldots, \alpha_{N-1})'$ is determined by

$$\boldsymbol{G}\boldsymbol{\alpha} = \boldsymbol{\beta}\ ,\ \boldsymbol{G} := (\langle \boldsymbol{B}_k, \boldsymbol{B}_j \rangle)_{j,k=0}^{N-1}\ ,\ \boldsymbol{\beta} := (\langle \boldsymbol{A}_N, \boldsymbol{B}_j \rangle)_{j=0}^{N-1}\ .$$

Proof: 1. By (4.5), it follows for $\boldsymbol{M}_N := \boldsymbol{O}_N' \, (\mathrm{diag}\, \boldsymbol{d}) \, \boldsymbol{O}_N \in \mathcal{A}_{O_N}$ that

$$||\boldsymbol{M}_N - \boldsymbol{A}_N||_F = ||\mathrm{diag}\, \boldsymbol{d} - \boldsymbol{O}_N \boldsymbol{A}_N \boldsymbol{O}_N'||_F$$

which implies (4.6) by the definition of the optimal preconditioner.
2. The computation of the optimal preconditioner of $\boldsymbol{A}_N$ in $\mathcal{A}_{O_N}$ is equivalent with the computation of the element of best approximation of $\boldsymbol{A}_N$ in the linear subspace $\mathcal{A}_{O_N}$ of the Hilbert space $\mathbb{R}^{N,N}$ equipped with the inner product (4.4). This can be done by the Galerkin approach (4.7). ∎

Now it is easy to verify that an optimal preconditioner of $\boldsymbol{T}_N' \boldsymbol{T}_N$ satisfies property (P1).

Corollary 4.2. Let $\boldsymbol{A}_N \in R^{N,N}$ be a symmetric positive definite matrix and let $\mathcal{A}_{O_N}$ be defined by (4.2) with respect to some orthogonal matrix $\boldsymbol{O}_N$. Then the optimal preconditioner $\boldsymbol{M}_N = \boldsymbol{M}_N(\boldsymbol{A}_N)$ is also symmetric and positive definite.

Proof: The symmetry of $\boldsymbol{M}_N$ follows by definition of $\mathcal{A}_{O_N}$.
By (4.6), the eigenvalues of $\boldsymbol{M}_N$ are given by the diagonal entries $(\boldsymbol{O}_N\boldsymbol{A}_N\boldsymbol{O}_N')_{k,k}$ $(k = 0,\ldots,N-1)$ of $\boldsymbol{O}_N\boldsymbol{A}_N\boldsymbol{O}_N'$. Since $\boldsymbol{A}_N$ is positive definite, it holds that

$$\begin{aligned} 0 &< \min\left\{\frac{\boldsymbol{x}_N'\boldsymbol{A}_N\boldsymbol{x}_N}{\boldsymbol{x}_N'\boldsymbol{x}_N} : \boldsymbol{x}_N \neq \boldsymbol{0}_N\right\} = \min\left\{\frac{\boldsymbol{y}_N'\boldsymbol{O}_N\boldsymbol{A}_N\boldsymbol{O}_N'\boldsymbol{y}_N}{\boldsymbol{y}_N'\boldsymbol{y}_N} : \boldsymbol{y}_N \neq \boldsymbol{0}_N\right\} \\ &\leq \boldsymbol{e}_k'(\boldsymbol{O}_N\boldsymbol{A}_N\boldsymbol{O}_N')\boldsymbol{e}_k = (\boldsymbol{O}_N\boldsymbol{A}_N\boldsymbol{O}_N')_{k,k} \quad (k = 0,\ldots,N-1)\,, \end{aligned}$$

and we are done. ■

In connection with (4.6) and Corollary 4.2 see also [18].

5 Clusters of eigenvalues

Let $C_{2\pi}$ denote the Banach space of 2π-periodic complex-valued functions equipped with the usual norm $\|\cdot\|_\infty$. In this section, we are only interested in functions $f = f_R + if_I \in C_{2\pi}$ with real Fourier coefficients

$$t_k := \int_{-\pi}^{\pi} f(x)e^{-ikx}\,\mathrm{d}x \quad (k \in \mathbb{Z})\,,$$

where we suppose that the real and the imaginary part of f

$$\begin{aligned} f_R &= t_0 + \sum_{k=1}^{\infty}(t_k + t_{-k})\cos kx\,, \\ f_I &= \sum_{k=1}^{\infty}(t_k - t_{-k})\sin kx \end{aligned}$$

do not vanish, respectively. Moreover, we assume that $\{t_k\}_{k\in\mathbb{Z}} \in l_1$, such that f belongs to the Wiener class. Consider the *N-th Toeplitz matrix corresponding to the generating function f*

$$\boldsymbol{T}_N := \mathrm{toeplitz}((t_0, t_{-1}, \ldots, t_{-(N-1)}), (t_0, t_1, \ldots, t_{N-1}))\,.$$

It is well-known that the singular values of T_N are distributed as $|f|$ [20]. Note that the above result was extended to functions $f \in L^2_{2\pi} \supset C_{2\pi}$ in [22].
The following definition is due to E.E. Tyrtyshnikov [22]. Let $\{\sigma_k^N\}_{k=1}^N$ be a sequence of real numbers and let $\gamma_N(\varepsilon)$ denote the number of those among σ_k^N $(k = 1,\ldots,N)$ which are outside the ε–ball centered at p. If $\gamma_N(\varepsilon) < K(\varepsilon)$, where $K(\varepsilon)$ is independent of N, then p is called a *proper cluster*. In this sense, we say that the values σ_k^N are clustered at p.
In the following, we restrict our attention to preconditioners of $\mathcal{A}_{C_N^{II}}$. By Theorem 3.1, the approach for the preconditioners with respect to the DST–I, the DST–II, the

DST-IV and the DCT-IV follows the same lines. Let $\boldsymbol{C} = \boldsymbol{C}_N := \boldsymbol{Z}_{N,2}\boldsymbol{C}_N^{II}$ and $\boldsymbol{S} = \boldsymbol{S}_N := \boldsymbol{Z}_{N,1}\boldsymbol{S}_N^{II}$. Regarding Theorem 3.1 ii), we associate with the sequence $\{\boldsymbol{T}_N\}_{N\in\mathbb{N}}$ of Toeplitz matrices a sequence $\{\boldsymbol{H}_N\}_{N\in\mathbb{N}}$ of Hankel matrices

$$\boldsymbol{H}_N := \text{hankel}((t_{-1},\ldots,t_{-(N-1)},0),(t_1,\ldots,t_{N-1},0))\,.$$

Let $\boldsymbol{M}_N$ denote the optimal preconditioner of $\boldsymbol{T}_N'\boldsymbol{T}_N$ with respect to the DCT–II, i.e.

$$\boldsymbol{M}_N := (\boldsymbol{C}_N^{II})'\,\delta(\boldsymbol{C}_N^{II}\,\boldsymbol{T}_N'\boldsymbol{T}_N\,(\boldsymbol{C}_N^{II})')\,\boldsymbol{C}_N^{II}\,. \tag{5.1}$$

In this section, we prove that under certain assumptions on $\boldsymbol{T}_N$, the eigenvalues of the preconditioned matrix $\boldsymbol{M}_N^{-1}\boldsymbol{T}_N'\boldsymbol{T}_N$ are clustered at 1. We follow the lines of R.H. Chan. First, we show that for all $\varepsilon > 0$ and N sufficiently large, the matrix $\boldsymbol{T}_N'\boldsymbol{T}_N - \boldsymbol{M}_N$ splits into a matrix of low rank independent of N and a matrix with spectral norm smaller than ε. Then we apply Cauchy's interlace theorem to verify that the eigenvalues of

$$\boldsymbol{M}_N^{-1}(\boldsymbol{T}_N'\boldsymbol{T}_N - \boldsymbol{M}_N) = \boldsymbol{M}_N^{-1}\boldsymbol{T}_N'\boldsymbol{T}_N - \boldsymbol{I}_N$$

are clustered at 0. Again, we drop the index N, if the dimension of the matrices follows from the context.

In preparation of Theorem 5.3, we provide the following two lemmata.

Lemma 5.1. Let $a = \{a_k\}_{k=0}^{\infty} \in l_1$ and $b = \{b_k\}_{k=0}^{\infty} \in l_1$. Then, for all $\varepsilon > 0$, there exists $m = m(\varepsilon)$ such that for all $N \geq 2m$ the Hankel matrix

$$\boldsymbol{H} := \text{hankel}((a_0,a_1,\ldots,a_{N-1}+b_{N-1}),(b_0,b_1,\ldots,a_{N-1}+b_{N-1}))$$

splits as $\boldsymbol{H} = \boldsymbol{V}_H + \boldsymbol{W}_H$, where

$$\boldsymbol{W}_H := \text{hankel}((a_0,\ldots,a_{m-1},\boldsymbol{0}_{N-m}),(b_0,\ldots,b_{m-1},\boldsymbol{0}_{N-m}))$$

is a matrix of rank $\leq 2m$ and where $\boldsymbol{V}_H := \boldsymbol{H} - \boldsymbol{W}_H$ satisfies $||\boldsymbol{V}_H||_2 < \varepsilon$.

Proof: Since $a, b \in l_1$, there exists for all $\varepsilon > 0$ an integer $m = m(\varepsilon)$ such that

$$\sum_{k=m}^{\infty} |a_k| < \varepsilon/2\,,\ \sum_{k=m}^{\infty} |b_k| < \varepsilon/2\,.$$

Now the assertion follows from $||\boldsymbol{V}_H||_2 \leq ||\boldsymbol{V}_H||_1 < \varepsilon/2 + \varepsilon/2$. ∎

Lemma 5.2. Let $t = \{t_k\}_{k\in\mathbb{Z}} \in l_1$ with $||t||_1 = \tau$ be given. Further let $\boldsymbol{T} = \boldsymbol{T}_N$ and $\boldsymbol{H} = \boldsymbol{H}_N$ be the corresponding N-th Toeplitz matrix and N-th Hankel matrix, respectively. Then, for all $\varepsilon > 0$, there exists $m = m(\varepsilon)$ such that for all $N \geq 4m$

$$\boldsymbol{H}\boldsymbol{T} + \boldsymbol{T}'\boldsymbol{H} + \boldsymbol{H}^2 = \boldsymbol{V} + \boldsymbol{W}\,, \tag{5.2}$$

where $||\boldsymbol{V}||_2 < \varepsilon$ and

$$\boldsymbol{W} := (w_{j,k})_{j,k=0}^{N-1} \text{ with } w_{j,k} = 0 \text{ for } 2m \leq j+k \leq 2N-2-2m. \tag{5.3}$$

Proof: By construction, it holds that $||\boldsymbol{T}||_2 \leq \tau$, $||\boldsymbol{H}||_2 \leq \tau$. Since $t \in l_1$, there exists $m = m(\varepsilon)$ such that

$$\sum_{|k|=m+1}^{\infty} |t_k| < \frac{\varepsilon}{6\tau}. \tag{5.4}$$

Then

$$\boldsymbol{T} = \boldsymbol{T}_\varepsilon + \boldsymbol{T}_B \ , \ \boldsymbol{H} = \boldsymbol{H}_\varepsilon + \boldsymbol{H}_B \tag{5.5}$$

with

$$\begin{aligned}
\boldsymbol{T}_\varepsilon &:= \text{toeplitz}((\mathbf{0}_{m+1}, t_{-(m+1)}, \ldots, t_{-(N-1)}), (\mathbf{0}_{m+1}, t_{m+1}, \ldots, t_{N-1}))\,, \\
\boldsymbol{T}_B &:= \text{toeplitz}((t_0, \ldots, t_{-m}, \mathbf{0}_{N-m-1}), (t_0, \ldots, t_m, \mathbf{0}_{N-m-1}))\,, \\
\boldsymbol{H}_\varepsilon &:= \text{hankel}((\mathbf{0}_m, t_{-(m+1)}, \ldots, t_{-(N-2)}, 0), (\mathbf{0}_m, t_{m+1}, \ldots, t_{N-2}, 0))\,, \\
\boldsymbol{H}_B &:= \text{hankel}((t_{-1}, \ldots, t_{-m}, \mathbf{0}_{N-m}), (t_1, \ldots, t_m, \mathbf{0}_{N-m}))\,,
\end{aligned}$$

where we obtain by (5.4) that $||\boldsymbol{T}_\varepsilon||_2 \leq \frac{\varepsilon}{6\tau}$, $||\boldsymbol{H}_\varepsilon||_2 \leq \frac{\varepsilon}{6\tau}$. Substituting (5.5) in (5.2), we obtain the desired decomposition

$$\begin{aligned}
\boldsymbol{H}\boldsymbol{T} + \boldsymbol{T}'\boldsymbol{H} + \boldsymbol{H}^2 &= (\boldsymbol{H}_\varepsilon + \boldsymbol{H}_B)(\boldsymbol{T}_\varepsilon + \boldsymbol{T}_B) + (\boldsymbol{T}'_\varepsilon + \boldsymbol{T}'_B)(\boldsymbol{H}_\varepsilon + \boldsymbol{H}_B) + (\boldsymbol{H}_\varepsilon + \boldsymbol{H}_B)^2 \\
&= (\boldsymbol{H}\boldsymbol{T}_\varepsilon + \boldsymbol{T}'_\varepsilon\boldsymbol{H} + \boldsymbol{H}_\varepsilon\boldsymbol{T}_B + \boldsymbol{T}'_B\boldsymbol{H}_\varepsilon + \boldsymbol{H}_\varepsilon\boldsymbol{H} + \boldsymbol{H}_B\boldsymbol{H}_\varepsilon) \\
&\quad + (\boldsymbol{T}'_B\boldsymbol{H}_B + \boldsymbol{H}_B\boldsymbol{T}_B + \boldsymbol{H}_B^2)\,.
\end{aligned}$$

■

Theorem 5.3. Let $t = \{t_k\}_{k\in\mathbb{Z}} \in l_1$ with $||t||_1 = \tau$ be given and let $\boldsymbol{T} = \boldsymbol{T}_N$ be the corresponding N–th Toeplitz matrix. Then, for all $\varepsilon > 0$, there exists $m = m(\varepsilon)$ such that for all $N \geq 2m$

$$\boldsymbol{T}'\boldsymbol{T} = \boldsymbol{C}'\boldsymbol{D}\boldsymbol{C} + \boldsymbol{V} + \boldsymbol{W},$$

where $\boldsymbol{D}$ denotes some diagonal matrix,

$$\boldsymbol{W} := (w_{j,k})_{j,k=0}^{N-1} \text{ with } w_{j,k} = 0 \text{ for } m \leq j+k \leq 2N-2-m,$$

and where $||\boldsymbol{V}||_2 < \varepsilon$.

Proof: Let $\boldsymbol{H} = \boldsymbol{H}_N$ denote the N–Hankel matrix associated with t. Then it follows by Theorem 3.1 ii) that

$$\begin{aligned}
\boldsymbol{T}'\boldsymbol{T} &= (\boldsymbol{T}' + \boldsymbol{H} - \boldsymbol{H})(\boldsymbol{T} + \boldsymbol{H} - \boldsymbol{H}) \\
&= (\boldsymbol{C}'\boldsymbol{D}_a\boldsymbol{C} + \boldsymbol{S}'\tilde{\boldsymbol{D}}_b\boldsymbol{C} - \boldsymbol{H})(\boldsymbol{C}'\boldsymbol{D}_a\boldsymbol{C} + \boldsymbol{C}'\tilde{\boldsymbol{D}}_b\boldsymbol{S} - \boldsymbol{H})\,,
\end{aligned}$$

where

$$\begin{aligned} \boldsymbol{a} = \boldsymbol{a}_{N+1} &:= (2t_0, t_{-1} + t_1, \ldots, t_{-(N-1)} + t_{N-1}, 0)' \in \mathbb{R}^{N+1}, \\ \boldsymbol{b} = \boldsymbol{b}_{N-1} &:= (t_{-1} - t_1, \ldots, t_{-(N-1)} - t_{N-1})' \in \mathbb{R}^{N-1} \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{D}_a = \mathrm{diag}(d_0, \ldots, d_{N-1}, d_N) \quad &, \quad (d_0, \ldots, d_N)' := \tilde{\boldsymbol{C}}^I_{N+1} \boldsymbol{a}, \\ \tilde{\boldsymbol{D}}_b = \mathrm{diag}(0, \tilde{d}_1, \ldots, \tilde{d}_{N-1}, 0) \quad &, \quad (\tilde{d}_0, \ldots, d_{N-1})' := \tilde{\boldsymbol{S}}^I_{N-1} \boldsymbol{b}. \end{aligned}$$

By $\boldsymbol{C}^{II}_N (\boldsymbol{C}^{II}_N)' = \boldsymbol{I}_N$, we further obtain that

$$\boldsymbol{T}'\boldsymbol{T} = \boldsymbol{C}' \boldsymbol{D}_a^2 \boldsymbol{C} + \boldsymbol{S}' \tilde{\boldsymbol{D}}_b^2 \boldsymbol{S} + \boldsymbol{C}' \boldsymbol{D}_a \tilde{\boldsymbol{D}}_b \boldsymbol{S} + \boldsymbol{S}' \tilde{\boldsymbol{D}}_b \boldsymbol{D}_a \boldsymbol{C} - \boldsymbol{H}(\boldsymbol{T} + \boldsymbol{H}) - (\boldsymbol{T}' + \boldsymbol{H})\boldsymbol{H} + \boldsymbol{H}^2$$

and by Theorem 3.1 ii) that

$$\boldsymbol{T}'\boldsymbol{T} = \boldsymbol{C}' \boldsymbol{D}_a^2 \boldsymbol{C} + \boldsymbol{C}' \tilde{\boldsymbol{D}}_b^2 \boldsymbol{C} - \boldsymbol{H}_{\tilde{D}_b^2} + \tilde{\boldsymbol{H}}_{\tilde{D}_b D_a} - (\boldsymbol{H}\boldsymbol{T} + \boldsymbol{T}'\boldsymbol{H} + \boldsymbol{H}^2) \tag{5.6}$$

with

$$\begin{aligned} \boldsymbol{H}_{\tilde{D}_b^2} &:= \mathrm{shankel}(u_1, \ldots, u_N), \\ (u_0, \ldots, u_N)' &:= \frac{2}{N} \tilde{\boldsymbol{C}}^I_{N+1} (0, \tilde{d}_1^2, \ldots, \tilde{d}_{N-1}^2, 0)', \\ \tilde{\boldsymbol{H}}_{\tilde{D}_b D_a} &:= \mathrm{ahankel}(v_1, \ldots, v_{N-1}, 0), \\ (v_1, \ldots, v_{N-1})' &:= \frac{2}{N} \tilde{\boldsymbol{S}}^I_{N-1} (d_1 \tilde{d}_1, \ldots, d_{N-1} \tilde{d}_{N-1})'. \end{aligned}$$

Set

$$\hat{\boldsymbol{H}} := \boldsymbol{H}_{\tilde{D}_b^2} - \tilde{\boldsymbol{H}}_{\tilde{D}_b D_a} = \mathrm{hankel}((w_1, \ldots, w_N), (w_{-1}, \ldots, w_{-N})) \tag{5.7}$$

with $w_N = w_{-N} = u_N$, $w_k = u_k - v_k$, $w_{-k} = u_k + v_k$ $(k = 1, \ldots, N-1)$. Then we have for all $N \in \mathbb{N}$ that

$$\begin{aligned} w_N &= \frac{2}{N} \sum_{k=1}^{N-1} (-1)^k \tilde{d}_k^2, \\ |w_N| &\leq \frac{2}{N} \|\tilde{\boldsymbol{S}}^I_{N-1} \boldsymbol{b}\|_2^2 \leq \|\boldsymbol{b}\|_2^2 \leq \|\boldsymbol{b}\|_1^2 \leq \tau^2. \end{aligned} \tag{5.8}$$

Moreover, we get by Theorem 3.1 i) for the first row $\boldsymbol{w} := (w_1, \ldots, w_{N-1})'$ of $\hat{\boldsymbol{H}}$ that

$$\begin{aligned} \boldsymbol{w} &= \frac{2}{N} \boldsymbol{R}'_N \tilde{\boldsymbol{C}}^I_{N+1} \mathrm{diag}(0, \tilde{d}_1, \ldots, \tilde{d}_{N-1}, 0) \boldsymbol{R}_N \tilde{\boldsymbol{S}}^I_{N-1} \boldsymbol{b} \\ &\quad - \frac{2}{N} \tilde{\boldsymbol{S}}^I_{N-1} \boldsymbol{R}'_N \mathrm{diag}(d_0, \ldots, d_N) \boldsymbol{R}_N \tilde{\boldsymbol{S}}^I_{N-1} \boldsymbol{b} \\ &= \mathrm{toeplitz}\left((-t_0, -t_1, \ldots, -t_{N-2}), (-t_0, -t_{-1}, \ldots, -t_{-(N-2)})\right) \boldsymbol{b} \\ &\quad + \mathrm{hankel}\left((t_{-2}, \ldots, t_{-(N-2)}, b_{N-1}, 0), (t_2, \ldots, t_{N-2}, -b_{N-1}, 0)\right) \boldsymbol{b}. \end{aligned}$$

Thus, it holds for all $N \in \mathbb{N}$ that

$$||\boldsymbol{w}||_1 \leq 2\,\tau^2\,.$$

Similarly, we conclude that

$$||(w_{-1},\ldots,w_{-(N-1)})'||_1 \leq \tau^2\,.$$

Together with (5.8), we see that $\hat{\boldsymbol{H}}$ satisfies the assertion of Lemma 5.1. Hence, for fixed $\varepsilon > 0$, there exists $\hat{m} = \hat{m}(\varepsilon) > 0$ such that for all $N \geq 2\hat{m}$

$$\hat{\boldsymbol{H}} = \hat{\boldsymbol{V}} + \hat{\boldsymbol{W}} \tag{5.9}$$

with $||\hat{\boldsymbol{V}}||_2 \leq \varepsilon/2$, $\hat{\boldsymbol{W}} = \text{hankel}((w_1,\ldots,w_m,\boldsymbol{0}_{N-m}),(w_{-1},\ldots,w_{-m},\boldsymbol{0}_{N-m}))$.
Furthermore, by Lemma 5.2, there exists $\tilde{m} = \tilde{m}(\varepsilon) > 0$ such that for N sufficiently large

$$(\boldsymbol{HT} + \boldsymbol{T}'\boldsymbol{H} + \boldsymbol{H}^2) = \tilde{\boldsymbol{V}} + \tilde{\boldsymbol{W}} \tag{5.10}$$

with $||\tilde{\boldsymbol{V}}||_2 \leq \varepsilon/2$ and with a low rank matrix $\tilde{\boldsymbol{W}}$ of the form (5.3). Applying (5.9) and (5.10) in (5.6), we obtain the assertion

$$\boldsymbol{T}'\boldsymbol{T} = \boldsymbol{C}'(\boldsymbol{D}_a^2 + \tilde{\boldsymbol{D}}_b^2)\boldsymbol{C} + (\tilde{\boldsymbol{V}} - \hat{\boldsymbol{V}}) + (\tilde{\boldsymbol{W}}) - \hat{\boldsymbol{W}})$$

with $m := \max\{\hat{m}, 2\tilde{m}\}$. ∎

Lemma 5.4. For $m > 0$ and $N > 2m$, let $\boldsymbol{V} \in \mathbb{R}^{N,N}$ with $||\boldsymbol{V}||_2 < \varepsilon/2$ and

$$\boldsymbol{W} := (w_{j,k})_{j,k=0}^{N-1} \text{ with } w_{j,k} = 0 \text{ for } m \leq j+k \leq 2N-2-m,$$

be given. Set $\omega := \sum_{j,k=0}^{N-1} |w_{j,k}|$. Then it holds for $N > 4\omega/\varepsilon$ that

$$||\delta(\boldsymbol{C}_N^{II}\,(\boldsymbol{V}+\boldsymbol{W})\,(\boldsymbol{C}_N^{II})')||_2 < \varepsilon\,.$$

Note that for fixed m, the value ω does not depend on N.

Proof: On the one hand, we obtain that

$$||\delta(\boldsymbol{C}_N^{II}\,\boldsymbol{V}\,(\boldsymbol{C}_N^{II})')||_2 \leq ||\boldsymbol{C}_N^{II}\,\boldsymbol{V}\,(\boldsymbol{C}_N^{II})'||_2 = ||\boldsymbol{V}||_2 < \varepsilon/2\,,$$

and on the other hand that

$$\begin{aligned} ||\delta(\boldsymbol{C}_N^{II}\,\boldsymbol{W}\,(\boldsymbol{C}_N^{II})')||_2 &\leq \max_{n=0,\ldots,N-1} |\frac{2}{N}\,(\varepsilon_n^N)^2 \sum_{j,k=0}^{N-1} w_{j,k} \cos\frac{n(2j+1)\pi}{2N} \cos\frac{n(2k+1)\pi}{2N}| \\ &\leq 2\omega/N < \varepsilon/2 \quad (N > 4\omega/\varepsilon)\,. \end{aligned}$$

Now summation implies the assertion. ∎

Theorem 5.5. Let $t = \{t_k\}_{k\in\mathbb{Z}} \in l_1$ with $||t||_1 = \tau$ be given and let $\boldsymbol{T}_N$ be the corresponding N–th Toeplitz matrix. Moreover, assume that the singular values of $\boldsymbol{T}_N$ are larger than $\gamma > 0$ for all $N \in \mathbb{N}$. Let $\boldsymbol{M}_N = \boldsymbol{M}_N(\boldsymbol{T}'_N\boldsymbol{T}_N)$ denote the optimal preconditioner of $\boldsymbol{T}'_N\boldsymbol{T}_N$ in $\mathcal{A}_{C_N^{II}}$. Then the eigenvalues of $\boldsymbol{M}_N^{-1}\boldsymbol{T}'_N\boldsymbol{T}_N$ are clustered at 1.

Proof: By Corollary 4.2, it holds that $||\boldsymbol{M}_N^{-1}||_2 < 1/\gamma$ for all $N \in \mathbb{N}$. Let $\varepsilon > 0$ be fixed. Then we obtain by (5.1), Theorem 5.3 and Lemma 5.4 that for N sufficiently large, there exists $M \in \mathbb{N}$ independent of N such that

$$\boldsymbol{T}'_N\boldsymbol{T}_N - \boldsymbol{M}_N = \boldsymbol{V}_N + \boldsymbol{W}_N - (\boldsymbol{C}_N^{II})'\,\delta(\boldsymbol{C}_N^{II}\,(\boldsymbol{V}_N + \boldsymbol{W}_N)\,(\boldsymbol{C}_N^{II})')\,\boldsymbol{C}_N^{II} = \boldsymbol{U}_N + \boldsymbol{W}_N\,,$$

where $\boldsymbol{W}_N$ is of low rank M and where $||\boldsymbol{U}_N||_2 < \varepsilon\gamma$. Now

$$\begin{aligned}\boldsymbol{M}_N^{-1/2}\,\boldsymbol{T}'_N\boldsymbol{T}_N\,\boldsymbol{M}_N^{-1/2} - \boldsymbol{I}_N &= \boldsymbol{M}_N^{-1/2}\,\boldsymbol{U}_N\,\boldsymbol{M}_N^{-1/2} + \boldsymbol{M}_N^{-1/2}\,\tilde{\boldsymbol{W}}_N\,\boldsymbol{M}_N^{-1/2}\\ &= \tilde{\boldsymbol{V}}_N + \tilde{\boldsymbol{W}}_N\,,\end{aligned}$$

where $||\tilde{\boldsymbol{V}}_N||_2 \le \varepsilon$ and where the rank of $\tilde{\boldsymbol{W}}_N$ is at most M. Since $\tilde{\boldsymbol{V}}_N$ and $\tilde{\boldsymbol{W}}_N$ are symmetric matrices, we can apply Cauchy's interlace theorem [25], which implies that for N sufficiently large, at most M eigenvalues of $\tilde{\boldsymbol{V}}_N + \tilde{\boldsymbol{W}}_N$ have absolute value greater than ε. Now the assertion follows since $\boldsymbol{M}_N^{-1/2}\,\boldsymbol{T}'_N\boldsymbol{T}_N\,\boldsymbol{M}_N^{-1/2}$ and $\boldsymbol{M}_N^{-1}\,\boldsymbol{T}'_N\boldsymbol{T}_N$ possess the same eigenvalues. ∎

Remark. Under the above assumptions on $\boldsymbol{T}_N$ it was proved that the eigenvalues of $(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)^*(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)$, where $\boldsymbol{M}_N = \boldsymbol{C}_N(\boldsymbol{T}_N)$ denotes the optimal circulant preconditioner of $\boldsymbol{T}_N$, are clustered at 1 [8, 22]. In general, the eigenvalues of $(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)'(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)$ are not clustered at 1, if $\boldsymbol{M}_N$ is the optimal trigonometric preconditioner of $\boldsymbol{T}_N$. If $\boldsymbol{T}_N = (t_{j-k})_{j,k=0}^{N-1}$ with $t_0 = 1$ and $t_k = -t_{-k}$ $(k = 1, \ldots, N-1)$, then the optimal trigonometric preconditioners of $\boldsymbol{T}_N$ are $\boldsymbol{M}_N = \boldsymbol{I}_N$, i.e. we have no preconditioning. If, for example, $t_{-1} = -t_1 = 2$ and $t_k = 0$ $(|k| > 1)$, then the matrices $\boldsymbol{T}_{2N+1}$ $(N \in \mathbb{N})$ satisfy the assumption of Theorem 5.5. However, the eigenvalues of $\boldsymbol{T}'_{2N+1}\boldsymbol{T}_{2N+1}$ are given by $9 - 8\cos(j\pi/(N+1))$ $(j = 0, \ldots, N)$.

6 Construction of optimal preconditioners of $T'T$

In this section, we explain how optimal trigonometric preconditoners of $\boldsymbol{T}'_N\boldsymbol{T}_N$ can be constructed with $O(N\log N)$ arithmetical operations. In contrast to the construction of optimal trigonometric preconditioners of $\boldsymbol{T}_N$, we are confronted with the fact that $\boldsymbol{T}'_N\boldsymbol{T}_N$ is not a Toeplitz matrix. Again, we consider only $\boldsymbol{C}_N^{II}$–preconditioners. The approach for the DST–I, DST–II, DST–IV and DCT–IV follows the same lines.
For the construction of the optimal preconditioner $\boldsymbol{M}_N = \boldsymbol{M}_N(\boldsymbol{T}'_N\boldsymbol{T}_N)$ we use the representation (4.7) of $\boldsymbol{M}_N$ with the basis $\{\boldsymbol{B}_k^{II} : k = 0, \ldots, N-1\}$ of $\mathcal{A}_{C_N^{II}}$ [2], [13]:

$$\boldsymbol{B}_k^{II} := (C_N^{II})'\,\mathrm{diag}(U_k(c_l))_{l=0}^{N-1}\,\boldsymbol{C}_N^{II}\,,$$

where $c_l := \cos\frac{l\pi}{N}$ and where U_k denotes the k–th *Chebyshev polynomial of second kind*

$$U_k(x) \;:=\; \sin((k+1)\arccos x)/\sin(\arccos x) \quad (x \in (-1,1))\,.$$

Moreover, we apply that $\{\boldsymbol{B}_k^I : k = 0,\ldots,N-1\}$ with

$$\begin{aligned}\boldsymbol{B}_k^I \;&:=\; (\boldsymbol{C}_N^{II})'\,\mathrm{diag}(T_k(c_l))_{l=0}^{N-1}\,\boldsymbol{C}_N^{II}\\ &=\; \text{stoeplitz}\;\boldsymbol{e}_k' \;+\; \text{shankel}\;\boldsymbol{e}_{k-1}'\,,\quad (k=1,\ldots,N-1)\end{aligned} \tag{6.1}$$

$\boldsymbol{e}_{-1} := \boldsymbol{0}_N$, and with the k-th *Chebyshev polynomial of first kind*

$$T_k(x) \;:=\; \cos(k\arccos x) \quad (x \in [-1,1])\,,$$

is another basis of $\mathcal{A}_{C_N^{II}}$. Both bases are related by

$$\begin{aligned}\boldsymbol{B}_0^I \;&=\; \boldsymbol{B}_0^{II} = \boldsymbol{I}\,,\; \boldsymbol{B}_1^{II} = 2\boldsymbol{B}_1^I\,,\\ \boldsymbol{B}_j^{II} \;&=\; \boldsymbol{B}_{j-2}^{II} \;+\; 2\boldsymbol{B}_j^I \quad (j=2,\ldots,N-1)\,,\end{aligned} \tag{6.2}$$

where the last equation follows by $U_j = U_{j-2} + 2T_j$. Now we have by (4.7) that

$$\boldsymbol{M}_N \;=\; \sum_{k=0}^{N-1} \alpha_k\,\boldsymbol{B}_k^{II}$$

with

$$\boldsymbol{\alpha} = \boldsymbol{G}^{-1}\,\boldsymbol{\beta}^{II}\,,\; \boldsymbol{G} := \left(\langle \boldsymbol{B}_j^{II}, \boldsymbol{B}_k^{II}\rangle\right)_{j,k=0}^{N-1}\,,\; \boldsymbol{\beta}^{II} := \left(\langle \boldsymbol{T}_N'\,\boldsymbol{T}_N, \boldsymbol{B}_j^{II}\rangle\right)_{j=0}^{N-1}\,.$$

Clearly, we are not interested in $\boldsymbol{M}_N$ itself, but in the diagonal matrix $\mathrm{diag}\,\boldsymbol{d}$ with

$$\boldsymbol{M}_N = (\boldsymbol{C}_N^{II})'\,(\mathrm{diag}\,\boldsymbol{d})\,\boldsymbol{C}_N^{II}\,. \tag{6.3}$$

If $\boldsymbol{\alpha}$ is known, then we obtain $\mathrm{diag}\,\boldsymbol{d}$ by

$$\mathrm{diag}\,\boldsymbol{d} = \sum_{k=0}^{N-1} \alpha_k\,\boldsymbol{C}_N^{II}\,\boldsymbol{B}_k^{II}\,(\boldsymbol{C}_N^{II})' = \sum_{k=0}^{N-1} \alpha_k\,\mathrm{diag}\,(U_k(c_l))_{l=0}^{N-1}\,,$$

i.e. by definition of U_k by

$$d_0 \;=\; \sum_{k=0}^{N-1} (k+1)\,\alpha_k\,,\; d_k = \frac{\hat{\alpha}_k}{\sin\frac{k\pi}{N}} \quad (k=1,\ldots,N-1)\,, \tag{6.4}$$

$$(\hat{\alpha}_k)_{k=1}^{N-1} \;=\; \tilde{\boldsymbol{S}}_{N-1}^I\,(\alpha_{k-1})_{k=1}^{N-1}\,. \tag{6.5}$$

Thus the construction of $\boldsymbol{d}$ from given $\boldsymbol{\alpha}$ requires $O(N\log N)$ arithmetical operations. It remains to find an efficient construction of the coefficient vector $\boldsymbol{\alpha}$. Therefore we use the following lemmata.

Lemma 6.1. For $\boldsymbol{G} := \left(\langle \boldsymbol{B}_j^{II}, \boldsymbol{B}_k^{II} \rangle\right)_{j,k=0}^{N-1}$, it holds that

$$\boldsymbol{G}^{-1} = \frac{1}{2N} \begin{pmatrix} 3 & 0 & -1 & & & & \\ 0 & 2 & 0 & -1 & & & \\ -1 & 0 & 2 & 0 & -1 & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & -1 & 0 & 2 & 0 & -1 \\ & & & -1 & 0 & 3 & -2 \\ & & & & -1 & -2 & \frac{3N-2}{N} \end{pmatrix},$$

such that the vector multiplication with $\boldsymbol{G}^{-1}$ can be computed with $O(N)$ arithmetical operations.

Proof: We show that $\boldsymbol{G}^{-1}\,\boldsymbol{G} = \boldsymbol{I}$. By definition, we have for the (k,l)–th entry of $\boldsymbol{G}$ that

$$\begin{aligned} g_{k,l} &= \langle \boldsymbol{B}_k^{II}\,,\, \boldsymbol{B}_l^{II} \rangle = \mathrm{tr}\left((\boldsymbol{B}_l^{II})'\,\boldsymbol{B}_k^{II}\right) \\ &= \sum_{j=0}^{N-1} U_l(c_j)\,U_k(c_j)\,. \end{aligned}$$

Then we get for $l = 2, \ldots, N-3$ that

$$2g_{k,l} - g_{k,l-2} - g_{k,l+2} = \sum_{j=0}^{N-1} U_k(c_j)\;\left(2U_l(c_j) - U_{l-2}(c_j) - U_{l+2}(c_j)\right)$$

and further since

$$2U_l(x) - U_{l-2}(x) - U_{l+2}(x) = 4(1-x^2)\,U_l(x)$$

and by definition of U_l that

$$2g_{k,l} - g_{k,l-2} - g_{k,l+2} = 4\sum_{j=1}^{N-1} \sin\frac{(k+1)j\pi}{N} \sin\frac{(l+1)j\pi}{N} = 2N\,\delta_{k,l}$$

$(k = 0, \ldots, N-1;\; l = 2, \ldots, N-3)$. Straightforward computation for $l = 0, 1, N-2, N-1$ completes the proof. ∎

Set

$$\boldsymbol{\beta}^I := \left(\langle \boldsymbol{T}'_N \boldsymbol{T}_N, \boldsymbol{B}_j^I \rangle\right)_{j=0}^{N-1}.$$

Then we obtain by the recurrence relation (6.2) that

$$\beta_0^{II} = \beta_0^I\,,\; \beta_1^{II} = 2\beta_1^I\,,\; \beta_k^{II} = \beta_{k-2}^{II} + \beta_k^I\;(k = 2, \ldots, N-1)\,. \tag{6.6}$$

Thus we can compute $\boldsymbol{\beta}^{II}$ from $\boldsymbol{\beta}^{I}$ with $O(N)$ additions. The following construction of $\boldsymbol{\beta}^{I}$ is based on an idea of E.E. Tyrtyshnikov [21]. We split $\boldsymbol{T} = \boldsymbol{T}_N = (t_{j-k})_{j,k=0}^{N-1}$ into a lower and an upper triangular Toeplitz matrix

$$\boldsymbol{T} = \boldsymbol{T}_L + \boldsymbol{T}_R$$

with diagonal entries $t_0/2$. Then we obtain that

$$\boldsymbol{\beta}_k^I = \langle \boldsymbol{T}'_L \boldsymbol{T}_L, \boldsymbol{B}_k^I \rangle + \langle \boldsymbol{T}'_R \boldsymbol{T}_R, \boldsymbol{B}_k^I \rangle + \langle \boldsymbol{T}'_L \boldsymbol{T}_R + \boldsymbol{T}'_R \boldsymbol{T}_L, \boldsymbol{B}_k^I \rangle \quad (k = 0, \ldots, N-1)\,. \tag{6.7}$$

We consider the summands on the right-hand side. The matrix $\boldsymbol{T}'_L \boldsymbol{T}_R + \boldsymbol{T}'_R \boldsymbol{T}_L$ is a symmetric Toeplitz matrix. The matrices $\boldsymbol{T}'_L \boldsymbol{T}_L$ and $\boldsymbol{T}'_R \boldsymbol{T}_R$ have lost their Toeplitz structure. For $\boldsymbol{A} \in \mathbb{R}^{N,N}$, we introduce the vectors $\boldsymbol{s}(\boldsymbol{A}) = (s_k(\boldsymbol{A}))_{k=0}^{N-1}$, $\boldsymbol{h}(\boldsymbol{A}) = (h_k(\boldsymbol{A}))_{k=0}^{N-1}$ and $\tilde{\boldsymbol{h}}(\boldsymbol{A}) = (\tilde{h}_k(\boldsymbol{A}))_{k=0}^{N-1}$ by

$$\begin{aligned}
s_k(\boldsymbol{A}) &:= \langle \boldsymbol{A}, \mathrm{stoeplitz}((\varepsilon_k^N)^{-2} \boldsymbol{e}_k) \rangle \quad (k = 0, \ldots, N-1)\,, \\
h_k(\boldsymbol{A}) &:= \langle \boldsymbol{A}, \mathrm{hankel}(\boldsymbol{e}_k, 0) \rangle \quad (k = 0, \ldots, N-2)\,, \\
\tilde{h}_k(\boldsymbol{A}) &:= \langle \boldsymbol{A}, \mathrm{hankel}(0, \boldsymbol{e}_k) \rangle \quad (k = 0, \ldots, N-2)\,.
\end{aligned}$$

Set $h_{-1} := 0$ and $\tilde{h}_{-1} := 0$. Then it follows by (6.1) that

$$\langle \boldsymbol{A}, \boldsymbol{B}_k^I \rangle = (\varepsilon_k^N)^2 s_k(\boldsymbol{A}) + h_{k-1}(\boldsymbol{A}) + \tilde{h}_{k-1}(\boldsymbol{A}) \quad (k = 0, \ldots, N-1)\,. \tag{6.8}$$

Lemma 6.2. Let $\boldsymbol{A} := \boldsymbol{T}'_L \boldsymbol{T}_R + \boldsymbol{T}'_R \boldsymbol{T}_L$ and let $\boldsymbol{r} := \boldsymbol{T}'_R (t_0, t_1, \ldots, t_{N-1})'$. Then it holds that

$$\begin{aligned}
s_k(\boldsymbol{A}) &= 2(N-k) r_k \quad (k = 0, \ldots, N-1)\,, \\
h_0(\boldsymbol{A}) &= r_0\,,\ h_1(\boldsymbol{A}) = 2\, r_1\,, \\
h_k(\boldsymbol{A}) &= 2\, r_k \,+\, h_{k-2}(\boldsymbol{A}) \quad (k = 2, \ldots, N-2)\,, \\
\tilde{h}_k(\boldsymbol{A}) &= h_k(\boldsymbol{A}) \quad (k = 0, \ldots, N-2)\,.
\end{aligned}$$

Proof: Since $\boldsymbol{A} = \mathrm{stoeplitz}\, \boldsymbol{r}$, the assertion follows by definition of $\boldsymbol{s}, \boldsymbol{h}$ and $\tilde{\boldsymbol{h}}$. ∎

Lemma 6.3. Let $\boldsymbol{A} := \boldsymbol{T}'_R \boldsymbol{T}_R$ and let $\boldsymbol{r} := \left((\varepsilon_k^N)^2 t_{-k}\right)_{k=0}^{N-1}$. Then it holds that

$$\boldsymbol{s}(\boldsymbol{A}) = 2\, \mathrm{hankel}(N r_0, (N-1) r_1, \ldots, r_{N-1}), (0, \ldots, 0, r_{N-1}))\, \boldsymbol{r}\,, \tag{6.9}$$

$$\begin{aligned}
h_0(\boldsymbol{A}) &= x_0\,,\ h_1(\boldsymbol{A}) = x_1\,, \\
h_k(\boldsymbol{A}) &= h_{k-2}(\boldsymbol{A}) + x_k \quad (k = 2, \ldots, N-2)\,,
\end{aligned} \tag{6.10}$$

$$\begin{aligned}
\tilde{h}_0(\boldsymbol{A}) &= y_0/2\,,\ \tilde{h}_1(\boldsymbol{A}) = y_1\,, \\
\tilde{h}_k(\boldsymbol{A}) &= \tilde{h}_{k-2}(\boldsymbol{A}) + y_k \quad (k = 2, \ldots, N-2)\,,
\end{aligned} \tag{6.11}$$

where

$$\begin{aligned}
\boldsymbol{x} &:= \boldsymbol{T}'_R \boldsymbol{r}\,, \\
\boldsymbol{y} &:= \mathrm{hankel}(2(r_0, r_1, \ldots, r_{N-1}), (-r_2, -r_3, \ldots, -r_{N-1}, 0, 2r_{N-1}))\, \boldsymbol{r}\,.
\end{aligned}$$

Proof: Since $\boldsymbol{T}_R$ is an upper triangular Toeplitz matrix, it holds that

$$a_{j,k} = a_{j-1,k-1} + r_j r_k \quad (j,k = 1\ldots, N-1)\,. \tag{6.12}$$

Consequently, we obtain that

$$s_k = 2\sum_{j=0}^{N-k-1} a_{j+k,j} = \sum_{j=0}^{N-1-k} (N-j-k) r_{j+k} r_j$$

which yields (6.9). The recursions (6.10) and (6.11) follow by straightforward calculation from (6.12). ∎

Lemma 6.4. Let $\boldsymbol{B} := \boldsymbol{T}_R' \boldsymbol{T}_R$ and let $\boldsymbol{J} := \text{shankel}\, \boldsymbol{e}_{N-1}$ denote the N-th counteridentity. Then it holds that

$$\begin{aligned} s_k(\boldsymbol{B}) &= s_k(\boldsymbol{J}\,\boldsymbol{B}\,\boldsymbol{J}) \quad (k = 0,\ldots,N-1)\,, \\ h_k(\boldsymbol{B}) &= \tilde{h}_k(\boldsymbol{J}\,\boldsymbol{B}\,\boldsymbol{J}) \quad (k = 0,\ldots,N-2)\,, \\ \tilde{h}_k(\boldsymbol{B}) &= h_k(\boldsymbol{J}\,\boldsymbol{B}\,\boldsymbol{J}) \quad (k = 0,\ldots,N-2)\,, \end{aligned}$$

such that $\boldsymbol{s}(\boldsymbol{B}), \boldsymbol{h}(\boldsymbol{B})$ and $\tilde{\boldsymbol{h}}(\boldsymbol{B})$ can be computed by (6.9) – (6.11).

Proof: The relations for $\boldsymbol{s}(\boldsymbol{B}), \boldsymbol{h}(\boldsymbol{B})$ and $\tilde{\boldsymbol{h}}(\boldsymbol{B})$ follow by definition of $\boldsymbol{J}$. By

$$\boldsymbol{J}\,\boldsymbol{T}_L'\,\boldsymbol{T}_L\,\boldsymbol{J} = (\boldsymbol{J}\,\boldsymbol{T}_L\,\boldsymbol{J})'\,(\boldsymbol{J}\,\boldsymbol{T}_L\,\boldsymbol{J})$$

and since

$$\boldsymbol{J}\,\boldsymbol{T}_L\,\boldsymbol{J} = \text{toeplitz}((t_0,\ldots,t_{N-1}),(t_0,0,\ldots,0))$$

is an upper triangular Toeplitz matrix, we can calculate $\boldsymbol{s}(\boldsymbol{B}), \boldsymbol{h}(\boldsymbol{B})$ and $\boldsymbol{h}(\tilde{\boldsymbol{B}})$ by (6.9) – (6.11) with $\boldsymbol{A} = \boldsymbol{J}\,\boldsymbol{B}\,\boldsymbol{J}$ and $\boldsymbol{r} = \left((\varepsilon_k^N)^2 t_k\right)_{k=0}^{N-1}$. ∎

Theorem 6.5. Let $\boldsymbol{T}_N := (t_{j-k})_{j,k=0}^{N-1}$. Then the optimal preconditioner $\boldsymbol{M}_N \in \mathcal{A}_{C_N^{II}}$ of $\boldsymbol{T}_N'\boldsymbol{T}_N$ can be constructed with $O(N\log N)$ arithmetical operations.

Proof: We compute $\boldsymbol{\beta}^I$ by (6.7), (6.8) and by the Lemmata 6.2 – 6.4. Taking into account that the multiplication of a vector with a Toeplitz matrix or a Hankel matrix requires $O(N\log N)$ operations, the whole construction of $\boldsymbol{\beta}^I$ takes $O(N\log N)$ arithmetical operations. ¿From $\boldsymbol{\beta}^I$ we compute $\boldsymbol{\beta}^{II}$ by (6.6) with $O(N)$ additions. Using Lemma 6.1, we get $\boldsymbol{\alpha} := \boldsymbol{G}^{-1}\boldsymbol{\beta}^{II}$ at the cost of $O(N)$ arithmetical operations. Finally, the DST–I in (6.4) to obtain $\boldsymbol{d}$ from $\boldsymbol{\alpha}$ needs $O(N\log N)$ arithmetical operations and we are done with an arithmetical complexity of $O(N\log N)$. ∎

Remark. We can use similar ideas for the construction of the optimal trigonometric preconditioners with respect to the $\boldsymbol{C}_N^{IV}$, $\boldsymbol{S}_{N-1}^{I}$, $\boldsymbol{S}_N^{II}$ and $\boldsymbol{S}_N^{IV}$. For the corresponding bases of $\mathcal{A}_{O_N}$

$$\begin{aligned} \boldsymbol{B}_k^I &:= \boldsymbol{O}'_{\dim}\,\mathrm{diag}(T_k(c_l))_{l=0}^{\dim-1}\,\boldsymbol{O}_{\dim} \\ &= \text{stoeplitz}\,\boldsymbol{e}'_k + \text{hankel}(\boldsymbol{u}'_k,\,\boldsymbol{v}'_k)\,, \\ \boldsymbol{B}_k^{II} &:= \boldsymbol{O}'_{\dim}\,\mathrm{diag}(U_k(c_l))_{l=0}^{\dim-1}\,\boldsymbol{O}_{\dim} \end{aligned}$$

it holds that

$$\boldsymbol{G}^{-1} = \frac{1}{K}\begin{pmatrix} 3 & 0 & -1 & & & & \\ 0 & 2 & 0 & -1 & & & \\ -1 & 0 & 2 & 0 & -1 & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & -1 & 0 & 2 & 0 & -1 \\ & & & -1 & 0 & g_1 & g_2 \\ & & & & -1 & g_2 & g_3 \end{pmatrix},$$

where

$\boldsymbol{O}_{\dim}$	c_l	dim	$\boldsymbol{u}_k$	$\boldsymbol{v}_k$	K	g_1	g_2	g_3
$\boldsymbol{C}_N^{II}$	$\cos\frac{l\pi}{N}$	N	$\boldsymbol{e}_{k-1}$	$\boldsymbol{e}_{k-1}$	$2N$	3	-2	$\frac{3N-2}{N}$
$\boldsymbol{C}_N^{IV}$	$\cos\frac{(l+1)l\pi}{2N}$	N	$\boldsymbol{e}_{k-1}$	$-\boldsymbol{e}_{k-1}$	$2N$	1	0	1
$\boldsymbol{S}_{N-1}^{I}$	$\cos\frac{(l+1)\pi}{N}$	$N-1$	$-\boldsymbol{e}_{k-2}$	$-\boldsymbol{e}_{k-2}$	$2N+2$	2	0	3
$\boldsymbol{S}_N^{II}$	$\cos\frac{(l+1)\pi}{N}$	N	$-\boldsymbol{e}_{k-1}$	$-\boldsymbol{e}_{k-1}$	$2N$	3	2	$\frac{3N-2}{N}$
$\boldsymbol{S}_N^{IV}$	$\cos\frac{(2l+1)\pi}{2N}$	N	$-\boldsymbol{e}_{k-1}$	$\boldsymbol{e}_{k-1}$	$2N$	1	0	1

Note that the construction of the optimal preconditioner with respect to the $\boldsymbol{C}_N^{IV}$ or the $\boldsymbol{S}_N^{IV}$ is especially simple.

7 Numerical Results

Finally, we present four examples of nonsymmetric Toeplitz systems

$$\boldsymbol{T}_N\boldsymbol{x}_N = \boldsymbol{b}_N\,, \tag{7.1}$$

for which the preconditioning of the normal equation

$$\boldsymbol{T}'_N\boldsymbol{T}_N\boldsymbol{x}_N = \boldsymbol{T}'_N\boldsymbol{b}_N \tag{7.2}$$

by an optimal trigonometric preconditioner $\boldsymbol{M}_N = \boldsymbol{M}_N(\boldsymbol{T}'_N\boldsymbol{T}_N)$ of $\boldsymbol{T}'_N\boldsymbol{T}_N$ significantly accerelates the convergence of the CG–method. We refer to the CG–method applied to the normal equation (7.2) as NCG–method. The algorithms were realized for the optimal preconditioners with respect to the DCT–II, the DST–II, the DCT–IV and

the DST–IV, respectively. Clearly, we can also use the DST–I preconditioner. Note that for symmetric Toeplitz matrices, the DST–I preconditioned CG–method for the solution of (7.1) shows a similar convergence behaviour as the DST–II preconditioned CG–method.
The fast computation of the preconditioners in the initial step and the computation of the preconditioned NCG–method were implemented in Matlab and tested on a Sun SPARCstation 20. The fast trigonometric transforms appearing both in the initialization and in the NCG–steps were taken from the C–implementation based on [3] and [19] by using the cmex–programm.
As transform length we choose $N = 2^n$. The right–hand side $\boldsymbol{b}_N$ of (7.1) is the vector consisting of N ones. The preconditioned NCG–method starts with the zero vector and stops if $\|r^{(j)}\|_2/\|r^{(0)}\|_2 < 10^{-7}$, where $r^{(j)}$ denotes the residual vector after j iterations. Our test matrices are the following four Toeplitz matrices $\boldsymbol{T}_N = (t_{j-k})_{j,k=0}^{N-1}$:

i) (see [15])

$$t_n = \begin{cases} 1\backslash\log(2-n) & n \le 1\,, \\ 1\backslash\log(2-n) + 1\backslash(1+n) & n = 0\,, \\ 1\backslash(1+n) & n \ge 1 \end{cases}$$

ii) (see [15])

$$t_n = \begin{cases} 2 & n = 0\,, \\ -0.7\, t_{n+1} & n \le -1\,, \\ 0.9\, t_{n-1} & n \ge 1\,. \end{cases}$$

iii) Here we use the Toeplitz matrices $\boldsymbol{T}_N$ arising from the generating function

$$f(x) = x^2 \mathrm{e}^{\mathrm{i}x}\,.$$

iv)

$$t_n = \begin{cases} -1.5 & n = -1\,, \\ 2 & n = 0\,, \\ 0.5 & n = 1\,, \\ 0 & \text{else.} \end{cases}$$

As expected, also for large transform lenghts N, the initialisation and each NCG–step can be computed very fast which reflects the arithmetic complexity of $O(N \log N)$ for these computations. The four last columns of the following tables show the number of iterations required by the NCG–metod for the different trigonometric preconditioners. The second column contains the number of iteration steps of the NCG–method without preconditioning. The columns 3 and 4 contain the numbers of iterations required by the CG–method applied to the equation

$$(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)'(\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)\boldsymbol{x}_N = (\boldsymbol{M}_N^{-1}\boldsymbol{T}_N)\boldsymbol{M}_N^{-1}\boldsymbol{b}_N\,,$$

where $\boldsymbol{M}_N$ denotes the optimal preconditioner of $\boldsymbol{T}_N$ with respect to the DCT–II and the DST–II, respectively.

n	$\boldsymbol{I}_N$	DCT–II	DST–II	$\boldsymbol{C}_N^{II}$	$\boldsymbol{S}_N^{II}$	$\boldsymbol{C}_N^{IV}$	$\boldsymbol{S}_N^{IV}$
7	24	11	10	8	15	14	14
8	32	12	11	8	17	15	15
9	43	14	13	8	19	17	16
10	57	18	16	9	20	19	16
11	86	20	19	9	20	20	17
12	121	25	22	9	22	22	19
13	176	30	28	9	22	22	19

Table 1: Number of iterations for example i)

n	$\boldsymbol{I}_N$	DCT–II	DST–II	$\boldsymbol{C}_N^{II}$	$\boldsymbol{S}_N^{II}$	$\boldsymbol{C}_N^{IV}$	$\boldsymbol{S}_N^{IV}$
7	34	44	44	9	12	9	12
8	43	47	50	8	11	8	11
9	53	50	52	7	10	8	11
10	59	50	53	7	9	7	10
11	50	50	53	6	9	7	9
12	58	49	53	6	8	7	9
13	56	48	54	6	8	7	9

Table 2: Number of iterations for example ii)

n	$\boldsymbol{I}_N$	DCT–II	DST–I	$\boldsymbol{C}_N^{II}$	$\boldsymbol{S}_N^{II}$	$\boldsymbol{C}_N^{IV}$	$\boldsymbol{S}_N^{IV}$
5	84	72	68	29	21	47	37
6	311	124	176	52	26	84	74
7	1226	264	412	116	33	173	140
8	5220	626	980	256	40	405	302
9	**	1741	3341	664	74	1031	846

Table 3: Number of iterations for example iii)

n	$\boldsymbol{I}_N$	DCT–II	DST–II	$\boldsymbol{C}_N^{II}$	$\boldsymbol{S}_N^{II}$	$\boldsymbol{C}_N^{IV}$	$\boldsymbol{S}_N^{IV}$
6	88	21	37	21	9	25	24
7	201	31	67	27	8	31	30
8	435	45	125	36	8	39	38
9	929	66	294	47	9	72	65

Table 4: Number of iterations for example iv)

Although not all matrices in our examples fulfil the assumptions of Theorem 5.5, the preconditioning with an optimal trigonometric preconditioner of $\boldsymbol{T}_N'\boldsymbol{T}_N$ accerelates the convergence of the NCG–method significantly.
For all examples, the preconditioning with respect to the DCT–IV and the DST–IV leads to similar numbers of iteration steps. It is easy to check that for symmetric

Toeplitz matrices $\boldsymbol{T}_N$ the optimal preconditioners $\boldsymbol{M}_N(\boldsymbol{T}_N)$ with respect to the DCT–IV and DST–IV coincide.
Except of the second example, the number of iterations differs, if we apply the preconditioners with respect to the DCT–II and the DST–II. Heuristically, this can be explained by the different structures of $\mathcal{A}_{C_N^{II}}$ and $\mathcal{A}_{S_N^{II}}$ and how ,,good" our example matrices fit into this structure. A general criterion for the choice of the optimal trigonometric preconditioner would be interesting. In this direction, it is remarkable, that the optimal preconditioner $\boldsymbol{M}_N(\boldsymbol{T}_N) \in \mathcal{A}_{C_N^{II}}$ of the auto-covariance matrix $\boldsymbol{T}_N := (\rho^{|j-k|})_{j,k=0}^{N-1}$ is "asymptotically equivalent" to $\boldsymbol{T}_N$ if $N \to \infty$, $\rho \to 1$, while the optimal preconditioner $\boldsymbol{M}_N(\boldsymbol{T}_N) \in \mathcal{A}_{S_N^{II}}$ of $\boldsymbol{T}_N$ is "asymptotically equivalent" to $\boldsymbol{T}_N$ if $N \to \infty$, $\rho \to 0$ [16].

Acknowledgement. The authors wish to thank M. Tasche for his valueable comments concerning the relations between trigonometric transforms and Toeplitz–plus–Hankel matrices.

References

[1] O. Axelson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1996.

[2] E. Bozzo and C. Di Fiore, On the use of certain matrix algebras associated with discrete trigonometric transforms in matrix displacement decomposition, *SIAM J. Matrix Anal. Appl.* 16: 312–326 (1995).

[3] G. Baszenski and M. Tasche, Fast polynomial multiplication and convolution related to the discrete cosine transform, *Linear Alg. Appl.* (in print).

[4] E. Boman and I. Koltracht, Fast transform based preconditioners for Toeplitz equations, *SIAM J. Matrix Anal. Appl.* 16: 628–645 (1995).

[5] T.F. Chan, An optimal circulant preconditioner for Toeplitz systems, *SIAM J. Sci. Statist. Comput.* 9: 766–771 (1988).

[6] R.H. Chang and G. Strang, Toeplitz systems by conjugate gradients with circulant preconditioner, *SIAM J. Sci. Statist. Comput.* 10: 104–119 (1989).

[7] R.H. Chan and M.K. Ng, Conjugate gradient methods of Toeplitz systems, *SIAM Review* 38/3 (1996).

[8] R.H. Chan and M. Yeung, Circulant preconditioners for complex Toeplitz systems, *SIAM J. Numer. Anal.* 30: 1193–1207 (1993).

[9] R.H. Chan and M.K. Ng, Sine transform based preconditioners for symmetric Toeplitz systems, *Linear Alg. Appl.* 232: 237–259 (1996).

[10] R.H. Chan, T.F. Chan and C.K. Wong, Cosine transform based preconditioners for total variation minimization problems in image processing, Technical Report 95–8 (64), Department of Mathematics, The Chinese Universiy of Hong Kong, 1995.

[11] R.H. Chan, J.G Nagy and R.J Plemmons, Circulant preconditioned Toeplitz least square problems, *SIAM J. Matrix Anal. Appl.* 15: 80–97 (1992).

[12] R.H. Chan, J.G Nagy and R.J Plemmons, Displacement Preconditioners for Toeplitz Least Squares Iterations, *Elec. Trans. Numer. Anal.* 2: 44–56 (1994).

[13] G. Heinig, A. Bojanczyk, Transformation tecniques for Toeplitz and Toeplitz–plus–Hankel matrices, I. Transformations, *Linear Alg. Appl.* (submitted).

[14] G. Heinig and K. Rost, Representations of Toeplitz–plus–Hankel matrices using trigonometric transforms with application to fast matrix-vector multiplication, preprint, December 1996.

[15] T. Ku and C. Kuo, Spectral Properties of Preconditioned Rational Toeplitz Matrices: the Nonsymmetric Case, *SIAM J. Matrix Anal. Appl.* 14: 521–544 (1993).

[16] K.R. Rao and P. Yip, *Discrete Cosine Transform*, Academic Press, Boston, 1990.

[17] Y. Saad, Computational Fluid Dynamics, Lecture Series, van Karman Institute for Fluid Dynamics, 1994.

[18] V. Strela and E.E. Tyrtyshnikov, Some generalisations of circulant preconditioners, Matrix Methods and Algorithms, IVM RAN Moscow: 66–73, 1990.

[19] G. Steidl and M. Tasche, A polynomial approach to fast algorithms for discrete Fourier–cosine and Fourier–sine transforms, *Math. Comp.* 56: 281–296 (1991).

[20] S.V. Parter, On the distribution of singular values of Toeplitz matrices, *Lin. Alg. Appl.* 80: 115–130 (1986).

[21] E.E. Tyrtyshnikov, Optimal and superoptimal circulant preconditioners, *SIAM J. Matrix Anal. Appl.* 13: 459–473 (1992).

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

[23] E.E. Tyrtyshnikov, A.Y. Yeremin and N.L. Zamarashkin, Clusters - Preconditioners - Convergence, Preprint.

[24] Z. Wang, Fast algorithms for the discrete W transform and for the discrete Fourier transform, *IEEE Trans. Acoust. Speech Signal Process.* 32: 803–816 (1984).

[25] J. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.