# REVIVING CIRCULANT PRECONDITIONERS FOR ADAPTIVE MESH REFINEMENT *

K. WALL †

In collaboration with: James Adler, Xiaozhe Hu, Misha Kilmer

**Abstract.** We present a preconditioner for solving fractional partial differential equations (PDEs) on an adaptive mesh. Adaptive refinement of the problem domain results in a stiffness matrix with Toeplitz blocks along the main diagonal, while the fractional PDE yields a dense stiffness matrix, where off-diagonal blocks are stored as low-rank approximations. Our preconditioner utilizes ideas from the circulant preconditioner of Chan and Strang [SIAM Journal on Scientific Computing, 1989], which takes advantage of the Toeplitz blocks on the diagonal and also accounts for the low-rank nature of the off-diagonal blocks. We demonstrate its effectiveness at accelerating convergence for our systems and emphasize its efficient application. This work presents theoretical results about the spectral clustering of the preconditioned system. In order to prove these results, special consideration is taken on how the low-rank blocks perturb the eigenvalues of the Toeplitz block-diagonal system. Numerical tests for various fractional orders are used to inspect any assumptions and validate our results.

**Key words.** Preconditioner, Adaptive Refinement, Toeplitz, Circulant, DCT, DST

## 1. Introduction.

There has long been interest in solving Toeplitz linear systems efficiently. A matrix $A$ is called Toeplitz if $a_{ij} = a_{i-j}$, in other words, $A$ has constant diagonals. Arbitrary matrices have $\mathcal{O}(n^2)$ unique entries and are solved directly by traditional techniques in $\mathcal{O}(n^3)$ time, Since such a Toeplitz matrix has just $\mathcal{O}(n)$ unique entries, we may expect to be able to solve it in $\mathcal{O}(n^2)$ time. This is indeed the case via techniques such as Levinson's algorithm. TODO CITE. Even this improvement, however, is infeasible for sufficiently large systems. Instead we turn to iterative Krylov methods. For these methods we can still take advantage of Toeplitz structure by using circulant preconditioners.

A circulant matrix is a Toeplitz matrix, that additionally has the "wrap-around" property where the last entry each row is the first entry of the subsequent row.

$$A = \begin{pmatrix} a_0 & a_{-1} & \cdots & a_{-(n-1)} \\ a_1 & a_0 & \cdots & a_{-(n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 \end{pmatrix} \quad C = \begin{pmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & \cdots & c_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & \cdots & c_0 \end{pmatrix} \tag{1.1}$$

Equation 1: $A$ is a Toeplitz matrix and $C$ is circulant.

Toeplitz matrices commonly arise in PDE discretization, signal processing, and control theory. Often the Toeplitz matrices are also symmetric positive definite (SPD). Given an SPD Toeplitz system $Ax = b$, the idea introduced by Strang and Chan is to use certain circulant preconditioners $C$ so that $C^{-1}Ax = C^{-1}b$ is solved in fewer

*This work was funded by NSF.

†Tufts University, (kate.wall@tufts.edu, https://katejeanw.github.io/).

iterations [?].

We leverage this idea to build a preconditioner for stiffness matrices generated from the adaptive finite element for fractional PDEs. In this setting the problem is discretized on a nonuniform mesh, and the resulting stiffness matrix is dense. In the usual finite element setting an uniform mesh results in a Toeplitz stiffness matrix. In the adaptive setting, after an initial solve on a uniform grid, the error on each element is estimated and the elements with largest error are refined via bisection. It is often the case that neighboring elements are refined the same number of times. So although the mesh is not globally uniform, there are areas of local uniformity. To build an effective preconditioner, we will take advantage of these locally uniform areas and their corresponding Toeplitz blocks in the stiffness matrix.

citations for facts about FEM and Toeplitz matrices

Although dense, the stiffness matrix can be effectively stored as a hierarchical matrix ($\mathcal{H}$-matrix). Due to weaker interaction between elements that are further apart in the domain, off-diagonal blocks are well-suited for low rank approximation. (See [?] for more stiffness matrix details.) The low rank representation makes for fast computations, but complicates both the implementation of the preconditioner and the spectral clustering of the preconditioned system.

In this paper we investigate how to precondition such systems using circulant matrices. Our investigation is focused on $\mathcal{H}$-matrices as in [?], but the same methods could be used on any matrix with Toeplitz blocks on the diagonal. We prove the preconditioned system has eigenvalues clustered around 1 and demonstrate numerical results with superlinear convergence.

We emphasize that our unique contributions are:

- building circulant preconditioners for adaptive meshes
- proving the preconditioned system has eigenvalues clustered around 1
- something else? numerical results? dealing with low-rank blocks?

## 2. Background.

applications for fPDEs

One of the most common approaches to numerically solving PDEs is the finite element method (FEM). FEM requires the domain be broken into a grid or mesh. When each piece of the domain, or element, is the same size we say it is an uniform mesh.

Prove uniform mesh gives toeplitz here

If the mesh is not fine enough to give the desired accuracy the initial approach may be to increase $n$ and make each element smaller, keeping a uniform mesh. Alternatively, when different levels of granularity are required across the domain to achieve desired accuracy, adaptive meshes can be employed. This approach only increases the mesh size in certain subdomains. While the entire mesh is no longer uniform, each element is part of a locally uniform mesh.

toeplitz blocks

fractional PDEs mean all elements interact, though some weaker some stronger. Weak and strong interaction guiding a natural splitting. Good for low-rank approximation. Summary of matrix structure, spsd?

Toeplitz matrices have many unique properties that give rise to efficient algorithms (see for example [?]). For our purposes we focus on their connection to functions in the Weiener class. This will allow us to take our problem from matrix operator theory to function theory. Suppose we have a singly infinite, symmetric Toeplitz matrix,

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \\ a_1 & a_0 & a_1 & \ddots \\ a_2 & a_1 & a_0 & \ddots \\ & \ddots & \ddots & \ddots \end{bmatrix}.$$

Assume $\sum_{k=-\infty}^{\infty} |a_k| < \infty$. Then the function $a(z) = \sum_{k=-\infty}^{\infty} a_k z^k$ is real, positive, and in the Wiener class for $|z| = 1$. It will be convenient to define the corresponding truncated function for a finite subsection of the infinite matrix:

DEFINITION 2.1. *The $m \times m$ finite subsection of the singly infinite matrix $A$ is denoted $A_m$ and defined as*

$$A_m = \begin{bmatrix} a_0 & a_1 & \cdots & a_{m-1} \\ a_1 & a_0 & \cdots & a_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m-1} & a_{m-2} & \cdots & a_0 \end{bmatrix}.$$

*Similarly this matrix induces a function from the truncated series, $a_m(z) = \sum_{k=-m}^{m} a_k z^k$.*

boundaries is the problem always (not infinite, block division)

- FEM, adaptive mesh and GMG
- Hierarchical matrices
- Hankel
- circulant and DFT/FFT (displacement kernel's)
- discrete convolution for exact solution and the boundary problem

## 3. Preconditioner.

A good preconditioner for an iterative method must in general decrease the total number of iterations without increasing the cost of a single iteration. We borrow Kailath's [?] criteria for preconditioners, though similar criteria has been established since Bini and Benedetto [?]:

1. Complexity of constructing applying $\boldsymbol{\tau}$ should be $\mathcal{O}(m \log m)$.
2. A linear system with $\boldsymbol{\tau}$ should be solved in $\mathcal{O}(m \log m)$ operations.
3. The spectrum of $\boldsymbol{\tau}^{-1} A$ should be clustered around 1

We can make this last point more precise.

DEFINITION 3.1 (Eigenvalue Clustering). *For any $\varepsilon > 0$ we say the eigenvalues of a matrix $\boldsymbol{\tau}^{-1} A_m$ are clustered around 1 if there exists $N_1$ and $N_2$ such that for all $m > N_1$ there are at most $N_2$ eigenvalues of $\boldsymbol{\tau}^{-1} A_m$ that do not lie within $[1-\varepsilon, 1+\varepsilon]$.*

The use of circulant preconditioners for Toeplitz systems originates from [?]. Kailath showed how both Strang and Chan type preconditioners come from the kernel of displacement operators, they further detail eight specific preconditioners for each form of the discrete sine and cosine transforms.

apply in n log n time

In our numerical results we use type TODO, discussed in [?]. Though the results could be generalized to all eight forms. An important fact that we will use later:

eigenvalues as points on function

$A$ is Toeplitz block, $H$ is "Hankel Correction"

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1} \\ a_1 & a_0 & a_1 & \cdots & a_{n-3} & a_{n-2} \\ a_2 & a_1 & a_0 & \cdots & a_{n-4} & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-2} & a_{n-3} & a_{n-4} & \cdots & a_0 & a_1 \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_1 & a_0 \end{bmatrix} H = \begin{bmatrix} a_2 & a_3 & a_4 & \cdots & a_{n-1} & 0 & 0 \\ a_3 & a_4 & a_5 & \cdots & 0 & 0 & 0 \\ a_4 & a_5 & a_6 & \cdots & 0 & 0 & a_{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_5 & a_4 & a_3 \\ 0 & 0 & a_{n-1} & \cdots & a_4 & a_3 & a_2 \end{bmatrix}$$

$\boldsymbol{\tau} = A - H$

We'll also utilize the common assumptions that $a(z) = \sum_{k=-\infty}^{\infty} |a_k| < \infty$ and $a(z) > 0$ for each Toeplitz block.

To apply the $\tau$ preconditioner to our adaptive mesh we can think first of a block construction, (explicit) Identify Toeplitz blocks $A_j$, Build Hankel correction $H_j$, Construct little $\boldsymbol{\tau}_j = A_j - H_j$ . Assemble big $\mathcal{T} = \begin{bmatrix} \boldsymbol{\tau}_1 & & & \\ & \boldsymbol{\tau}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\tau}_m \end{bmatrix}$

To apply it $m \log m$ instead, (implicit) Get sizes of Toeplitz blocks $n_j$, calculate $\lambda_i = c_i[S\mathbf{t}]_i$ for $1 \leq i \leq n_j$, apply $S\Lambda^{-1}S$.

- story of many
- SPSD with SPSD A
- how it works w multigrid

**4. Theoretical Results.** Story: we know how this works on one block, what about block diagonal, what about with things happening outside of block diagonal?

**4.1. Eigenvalue Bounds.**

LEMMA 4.1 (Weyl's Inequality).
*Let $M$ and $E$ be Hermitian $n \times n$ matrices. Then for $A := M + E$ we have*

$$|\lambda_k(A) - \lambda_k(M)| \leq ||E||_2 \ , 1 \leq k \leq n.$$

*That is the eigenvalues of $A$ are at most $||E||_2$ away from the eigenvalues of $M$.*

**4.1.1. Kailath-Olshevsky Proof rewritten.** Fix $\varepsilon > 0$. Assumptions:

1. Generating function is in the Wiener Class,

$$a(z) = \sum_{k=-\infty}^{\infty} a_k z^k, \ \sum_{k=-\infty}^{\infty} |a_k| < \infty.$$

2. Generating function is bounded away from zero on unit circle,

$$a(z) > 2\varepsilon, \ |z| = 1.$$

LEMMA 4.2. *Let $a_m(z)$ be the truncated generating function with $2m-1$ terms, $\sum_{k=-(m-1)}^{m-1} a_k z^k$. Then for each $\lambda_k(S_Q(A))$ there exists $z_k$ on the unit circle such that $\lambda_k = a_m(z_k)$.*

LEMMA 4.3. *Choose $m$ and $N < m$ big enough so that $\sum_{N+1}^{\infty} |a_k| < \varepsilon$. Then assumption 2 implies that $a_m(z)$ is positive on the unit circle.*

*Proof.* First notice $\varepsilon > \sum_{N+1}^{\infty} |a_k| > \sum_{m}^{\infty} |a_k|$. Now

$$a(z) = a_m(z) + \sum_{-\infty}^{-m} a_k z^k + \sum_{m}^{\infty} a_k z^k$$

$$\implies a_m(z) = a(z) - \sum_{-\infty}^{-m} a_k z^k - \sum_{m}^{\infty} a_k z^k$$

$$\implies a(z) - 2\varepsilon < a_m(z)$$

$$0 < a(z) - 2\varepsilon < a_m(z)$$

∎

COROLLARY 4.4. *The matrices $S_Q(A_m)$ and $S_Q(A_m)^{-1}$ are positive definite.*

First we present the proof for a single Toeplitz matrix/block, adapted from [?]. This will be used in the proof of the full matrix spectral clustering.

Lemma statement

$$S_Q(A) = A + H + B \tag{4.1}$$

$$A = S_Q(A) - (H + B) \tag{4.2}$$

Where $A$ is Toeplitz (given), $H$ is Hankel, and $B$ is 'border' matrix, at most nonzero in exterior rows and columns. Thus

$$S_Q(A)^{-1}A = I - S_Q(A)^{-1}(H + B).$$

So it suffices to show the spectrum of $S_Q(A)^{-1}(H + B)$ is clustered around zero. Let $\varepsilon > 0$ and choose $N$ such that $\sum_{N+1}^{\infty} |a_k| < \varepsilon$. We can then split $H + B$ into the sum of a low-rank matrix $A_{lr}$ and a small norm matrix $A_{sn}$. Here $A_{lr}$ contains the diagonals with entries $a_0, \ldots, a_N$. Let $s := \operatorname{rank}(A_{lr}) << m$. Now $A_{sn} := (H + B) - A_{lr}$ is a hermitian $m \times m$ matrix with at most two copies of $a_{N+1}, \ldots, a_m$ in each row/column. Thus $||A_{sn}||_2 = \sqrt{||A_{sn}||_1 ||A_{sn}||_\infty} = ||A_{sn}||_1 < 2\varepsilon$. Hence by Weyl's Inequality at least $m - s$ of the eigenvalues of $H + B$ are clustered within $2\varepsilon$ of zero.

Now we use the min-max theorem to bound the eigenvalues of $S_Q(A)^{-1}(H + B)$.

$$\begin{aligned}
\lambda_k(S_Q(A)^{-1}(H+B)) &= \min_{\dim V = k} \max_{x \in V} \left( \frac{((H+B)x, x)}{(S_Q(A)x, x)} \right) \\
&\leq \min_{\dim V = k} \left[ \max_{x \in V} \left( \frac{((H+B)x, x)}{(x, x)} \right) \max_{x \in V} \left( \frac{(x, x)}{(S_Q(A)x, x)} \right) \right] \\
&\leq \left[ \min_{\dim V = k} \max_{x \in V} \left( \frac{((H+B)x, x)}{(x, x)} \right) \right] \max_{x \in \mathbb{R}^n} \left( \frac{(x, x)}{(S_Q(A)x, x)} \right) \\
&= \lambda_k(H+B) \max_{x \in \mathbb{R}^n} \left( \frac{(x, x)}{(S_Q(A)x, x)} \right) \\
&\leq \lambda_k(H+B) \frac{1}{\lambda_{\min}(S_Q(A))} \\
&= \lambda_k(H+B) \frac{1}{a_m(z_{\min})} \\
&\leq \lambda_k(H+B) \frac{1}{\min_{|z|=1} a_m(z)}
\end{aligned}$$

Can be simplified with $B = \mathbf{0}$. Actual condition: $a(z) > 2\varepsilon$.

### 4.2. Full Matrix Proof.

**4.2.1. setup.** A single block preconditioner is $\tau$ the block diagonal preconditioner is $T$.

On a single block we write $\tau = A - H$, but for the full adaptive matrix $A$ includes off diagonal blocks. Denote the diagonal (Toeplitz blocks) as $A_D$ and everything else $A_E$ so that

$$A = A_D + A_E + B$$

. And thus the splitting as in [**?**] is expressed $A_D = T + H$ and $A = A_E + B + T + H$. So

$$T^{-1}A = T^{-1}(T + H + A_E + B) = I + T^{-1}H + T^{-1}A_E + T^{-1}B \tag{4.3}$$

**4.2.2. Proof.** It suffices to show that $T^{-1}H$, $T^{-1}B$ and $T^{-1}A_E$ have spectra clustered around zero. First notice that $T^{-1}H$ is block diagonal and the spectrum of each block can be characterized using the former proof on each block.

since we don't really choose block size in practice the actual block size dictates the size of $\varepsilon$. Over all the blocks we can take the max $\varepsilon$ for a uniform bound, but many will be clustered tighter than that. Supports argument that bigger Toeplitz blocks = better clustering

Assume the off-diagonal-by-one blocks are low-rank. Let $C$ be such a block with dimensions $n_C \times n_C$ and rank $r_C << n_C$. Using the SVD we can split $C$ as

$$C = \left(\sum_{i=1}^{r_C} \sigma_i^{(C)} \mathbf{u}_i^{(C)} \mathbf{v}_i^{(C)}*\right) + \left(\sum_{i=r_C+1}^{n_C} \sigma_i^{(C)} \mathbf{u}_i^{(C)} \mathbf{v}_i^{(C)}*\right).$$

With a slight abuse of notation, we can embed this decomposition in the appropriate "off-diagonal" position of an $m \times m$ matrix. Doing this for all such off-diagonal blocks we write

$$\begin{aligned} B &= \sum_{C \in \text{off-diag}} \left[\left(\sum_{i=1}^{r_C} \sigma_i^{(C)} \mathbf{u}_i^{(C)} \mathbf{v}_i^{(C)}*\right) + \left(\sum_{i=r_C+1}^{n_C} \sigma_i^{(C)} \mathbf{u}_i^{(C)} \mathbf{v}_i^{(C)}*\right)\right] \\ &= \left(\sum_{i=1}^{r_B} \sigma_i \mathbf{u}_i \mathbf{v}_i*\right) + \left(\sum_{i=r_B+1}^{n_B} \sigma_i \mathbf{u}_i \mathbf{v}_i*\right) \end{aligned}$$

where $r_B = \max_{C \in \text{off-diag}} r_C$.

We additionally split $H$ by separating the anti-diagonals with coefficients $a_0, \dots, a_N$ and the anti-diagonals comprising of $a_{N+1}, \dots a_m$. So we have two splittings,

$$\begin{aligned} B &= \left(\sum_{i=1}^{r_B} \sigma_i \mathbf{u}_i \mathbf{v}_i*\right) + \left(\sum_{i=r_B+1}^{n_B} \sigma_i \mathbf{u}_i \mathbf{v}_i*\right) \\ H &= H|_{a_0,\dots,a_N} + H|_{a_{N+1},\dots,a_m}. \end{aligned}$$

The first term in each sum can be thought of as our 'low-rank' equivalent from before and similarly the second term is our 'small-norm' summand.

Bound on number of off diagonal blocks

Finally we can make the splitting $A = A_{SN} + A_{LR}$ where

$$A_{SN} = H|_{a_{N+1},\ldots,a_m} + \sum_{i=r_B+1}^{n_B} \sigma_i \mathbf{u}_i \mathbf{v}_i * + A_E$$

$$A_{LR} = H|_{a_0,\ldots,a_N} + \sum_{i=1}^{r_B} \sigma_i \mathbf{u}_i \mathbf{v}_i * .$$

$A_{LR}$ represent outliers, IE $s := \mathrm{rank}(A_{LR}) \leq N + r_B$ bounds the number of outliers.

Is the N part of this bound true? 2N?

So the work is showing $||T^{-1}A_{SN}||_2 \leq \varepsilon$. Define $\tilde{B} = \sum_{i=r_B+1}^{n_B} \sigma_i \mathbf{u}_i \mathbf{v}_i *$ and $\tilde{H} = H|_{a_{N+1},\ldots,a_m}$, so that $A_{SN} = \tilde{H} + \tilde{B} + A_E$.

$$||T^{-1}A_{SN}||_2 \leq ||T^{-1}\tilde{H}||_2 + ||T^{-1}\tilde{B}||_2 + ||T^{-1}A_E||_2$$

We can bound $||T^{-1}\tilde{H}||_2$ as in [?]. We can bound $||T^{-1}A_E||_2$ with Weyl's inequality:

$$||T^{-1}A_E||_2 \leq ||T^{-1}||_2||A_E||_2 = \sigma_{\max}(T^{-1})\sigma_{\max}(A_E) = \frac{\sigma_{\max}(A_E)}{\lambda_{\min}(T)}.$$

Finally we bound $||T^{-1}\tilde{B}||_2$.

$$\begin{aligned}
\lambda_k(T^{-1}\tilde{B}) &= \min_{\dim V = k} \max_{x \in V} \left( \frac{(\tilde{B}x, x)}{(Tx, x)} \right) \\
&\leq \min_{\dim V = k} \left[ \max_{x \in V} \left( \frac{(\tilde{B}x, x)}{(x, x)} \right) \max_{x \in V} \left( \frac{(x, x)}{(Tx, x)} \right) \right] \\
&\leq \left[ \min_{\dim V = k} \max_{x \in V} \left( \frac{(\tilde{B}x, x)}{(x, x)} \right) \right] \max_{x \in \mathbb{R}^n} \left( \frac{(x, x)}{(Tx, x)} \right) \\
&= \lambda_k(\tilde{B}) \max_{x \in \mathbb{R}^n} \left( \frac{(x, x)}{(Tx, x)} \right) \\
&\leq \lambda_k(\tilde{B}) \frac{1}{\lambda_{\min}(T)} \\
&= \lambda_k(\tilde{B}) \min_{n \in n_k} \min_{1 \leq i \leq n} \frac{\sin(\frac{\pi i}{n+1})}{\sum_{j=1}^{n} t_j \sin(\frac{\pi i j}{n+1})}
\end{aligned}$$

Since $\tilde{B}$ made of blocks that have form $\sum_{i=r_B+1}^{n_B} \sigma_i \mathbf{u}_i \mathbf{v}_i *$ what can we say about $\lambda_k$?

- numerical test confirming off-diag low rank
- Explanation and tests showing off-off-diag are small norm
- technically lots of $1 \times 1$ blocks at boundaries, these get jacobi inverse treatment so are clustered around 1
- Comment - all problems come from boundaries
- Extend proof to different kinds of circulant preconditioner

## 5. Numerical Results.

- enough info to reproduce
- Single block clustering
- Adaptive clustering (what happens to smallest eigenvalue?)

- behavior for different $\alpha$
- Verify assumptions from proof
- convergence of solving with PCG (superlinear convergence)

**6. Conclusion.** Future work: how to build adaptive mesh to increase block size, other circulant preconditioners, tensor preconditioners, higher dimension domain, mixed precision