Helsinki University of Technology Institute of Mathematics Research Reports
Teknillisen korkeakoulun matematiikan laitoksen tutkimusraporttisarja
Espoo 2003 A451

COMBINING NORMALITY WITH THE FFT TECHNIQUES

Marko Huhtanen

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Institute of Mathematics Research Reports
Teknillisen korkeakoulun matematiikan laitoksen tutkimusraporttisarja
Espoo 2003 A451

COMBINING NORMALITY WITH THE FFT TECHNIQUES

Marko Huhtanen

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics

Marko Huhtanen: *Combining normality with the FFT techniques* ; Helsinki University of Technology Institute of Mathematics Research Reports A451 (2003).

Abstract: *Ways to combine normality with the fast Fourier transformation ideas are studied by employing various matrix structures. The Toeplitz decomposition is natural for polynomially generating normal matrices while the so-called persymmetric splitting provides a framework for polynomially extending the Toeplitz matrix structure. In this context fast matrix-vector multiplications with the FFT techniques can be applied to different Toeplitz related matrices. Two sparse matrix methods for generating normal matrices are introduced to have more alternatives with normality. The method based on embedding matrices in normal matrices allows us to invert nonnormal matrices through inverting normal matrices. This is a potential approach for combining the FFT ideas with preconditioning nonnormal problems. To end with, we introduce a new iterative method.*

AMS subject classifications: 15A57, 65F10, 65T50

Keywords: normal matrix, FFT, Toeplitz matrix, persymmetric matrix, normal embedding, Kronecker product, preconditioning, 5-term recurrence

Marko.Huhtanen@hut.fi

ISBN 951-22-6152-9
ISSN 0784-3143
Inst. of Math. HUT, Espoo, 2002

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics
P.O. Box 1100, 02015 HUT, Finland
email:math@hut.fi http://www.math.hut.fi/

1 Introduction

A square matrix is perfectly suited to using iterative methods with if an optimal short-term recurrence can be executed with it and matrix-vector multiplications can be performed inexpensively. These are, admittedly, very stringent conditions to hold simultaneously. The first one renders normal matrices interesting because of the recent introduction of optimal methods for this particular class of matrices [22, 24, 9, 25, 27] while the second one calls, typically, for the FFT ideas in the dense matrix case. Of course, with circulant matrices we have both normality and the possibility to employ the FFT techniques so that there are matrices satisfying the prescribed two conditions. In this paper we consider other ways to combine normality with fast matrix-vector multiplication methods relying, typically, on the FFT techniques. Although this is a theoretical study with less emphasis on computations, these ideas can be used, for example, in preconditioning linear systems in a fashion similar to circulant matrices [37, 5].

We start by introducing matrix decompositions supporting both normality and the FFT ideas. The Toeplitz decomposition, i.e., when a matrix is split into its Hermitian and skew-Hermitian part, can be regarded as natural while dealing with normality [22, 24]. In particular, by forming polynomials in either of the parts, we always have a polynomial family of normal matrices. Recall that circulant matrices are also obtained through forming polynomials in a normal matrix. For a complete characterization of normal Toeplitz matrices, see [28].

The structure that supports employing the FFT techniques in a natural way is the so-called persymmetric splitting of a matrix. A matrix is persymmetric if it is symmetric with respect to the diagonal joining the left lower corner with the right upper corner. Every matrix can be uniquely split into a sum of a persymmetric and a skew-persymmetric matrix; see (6) for details. This decomposition is motivated by the simple remark that polynomials in Toeplitz matrices remain persymmetric. In this manner persymmetry provides a framework for generalizing the Toeplitz structure and thereby permits using the FFT ideas with a wider range of matrices. For a simple illustration, if $A = p(T)$ for a Toeplitz matrix $T \in \mathbb{C}^{n\times n}$ and a polynomial p, then after factoring the polynomial, matrix-vector products with A cost only of order $\deg(p)O(n \log n)$ operations. This is still very impressive (depending on the degree of the polynomial, of course), in particular, because every matrix is the product of two persymmetric matrices. If T is additionally normal, then algorithms for normal matrices can be executed.

To have more normal matrices readily available, two sparse matrix techniques are introduced to this end. We consider embedding matrices in normal matrices by taking any square matrix $B \in \mathbb{C}^{n\times n}$, a parameter $\lambda \in \mathbb{C}$, and setting

$$M \equiv M_{U,V,\lambda}(B) = \begin{bmatrix} B & U(B-\lambda I)^* \\ V(B-\lambda I)^* & B \end{bmatrix} \in \mathbb{C}^{2n\times 2n}, \tag{1}$$

with a pair of commuting unitary matrices U and V commuting with B.

Then M is normal (also called a normal dilation of B) and forming polynomials in M gives a family of normal matrices. Because of the particular block structure, this also supports the point of view of the persymmetric splitting. Moreover, to benefit from the Toeplitz ideas, natural choices for B are matrices with which matrix-vector multiplications can be performed fast.

The embedding (1) can also be employed in finding the inverse of a nonnormal matrix $A \in \mathbb{C}^{n\times n}$. This is based on extracting the $(1,1)$-block of the inverse of M which, of course, is never computed in practice. Only matrix-vector products with M are performed. This allows us to iteratively solve linear systems involving nonnormal matrices through solving linear systems involving normal matrices, without resorting to the normal equations. For finding a preconditioner for A with this approach, the problem boils down to choosing B, U, V and λ inexpensively. We introduce criteria to this end in case A is a Toeplitz matrix by using its symbol. These ideas are then extended to Toeplitz related problems and to persymmetric ones, in particular.

To have different block structures aside from (1), the other sparse matrix method for generating normal matrices relies on the Kronecker product through forming

$$p_{\otimes}(N_1, N_2) = \sum_{i,j} c_{i,j} N_1^i \otimes N_2^j \tag{2}$$

which is normal if the factors N_1 and N_2 are. Here $c_{i,j} \in \mathbb{C}$. In general, there seem to be more opportunities to apply the FFT ideas if the persymmetric part of a matrix dominates. The structure of $p_{\otimes}(N_1, N_2)$ allows us to use the FFT techniques with matrices having a dominating skew-persymmetric part.

To end with, we consider practical algorithms for the matrices introduced. We present an optimal method for solving linear systems involving normal matrices of type (1). Then, regardless of $\lambda \in \mathbb{C}$, by choosing $V = e^{i\alpha}U^*$, with $\alpha \in \mathbb{R}$, we can execute an optimal 5-term recurrence due to the fact that the spectrum of M is located on a second degree algebraic curve. The optimality condition obtained is quite impressive combining GMRES with that of the normal equations.

The paper is organized as follows. In section 2 we consider matrix decompositions supporting normality and the usage of the FFT techniques. We pay particular attention to persymmetry which yields various ways to regard a matrix as almost Toeplitz. In section 3 two methods for generating normal matrices are studied. The first one is based on embedding matrices in normal matrices and the second one on the Kronecker product. Section 4 deals with an optimal short-term recurrence.

2 Toeplitz related matrices supporting the FFT ideas

Regarding normality, forming polynomials in a normal matrix is a closed operation of which the set of circulant matrices is a classical example. The

Toeplitz decomposition of a square matrix $A \in \mathbb{C}^{n \times n}$, defined via

$$A = \frac{A + A^*}{2} + \frac{A - A^*}{2} = H + K, \tag{3}$$

supports this polynomial approach. Namely, A is normal if and only if its Hermitian part H and skew-Hermitian part K commute; see, e.g., [16, Condition 21]. In [22] this was employed by taking a Hermitian matrix $H \in \mathbb{C}^{n \times n}$ and a polynomial p with real coefficients and forming $H + ip(H)$. Then, by varying the Hermitian part and the polynomial, a dense subset of normal matrices is obtained. This construction is intrinsically complex and to deal with real matrices, one alternative is to employ the skew-Hermitian part instead.

Proposition 1 *The skew-Hermitian part of a real square matrix is generically nonderogatory.*

Proof. With $r = \lfloor \frac{n}{2} \rfloor$ and $U, V \in \mathbb{R}^{n \times r}$ consider the mapping

$$(U, V) \to f(U, V) := \sum_{k=1}^{r} (u_j v_j^* - v_j u_j^*) \tag{4}$$

to the set of real skew-Hermitian matrices. The image $f(U, V)$ is nonderogatory if and only if the dimension of the span of the columns of $[U\ V]$ is $2r$. □

Let $\mathcal{P}_e$ denote the set of polynomials with even powers.

Theorem 1 *The set of normal matrices $A \in \mathbb{R}^{n \times n}$ with a nonderogatory skew-Hermitian part K is of real dimension $\lceil \frac{n^2}{2} \rceil$. There exists a unique $p \in \mathcal{P}_e$ of degree $n - 1$ at most such that $A = p(K) + K$.*

Proof. The set of real skew-Hermitian nonderogatory matrices is of dimension $\frac{n^2 - n}{2}$. The eigenvalues of $A \in \mathbb{R}^{n \times n}$ are symmetrically located with respect to the real axis. If A is additionally normal with a nonderogatory skew-Hermitian part K, then with Lagrange interpolation we can find an interpolating polynomial p of degree $n - 1$ depending on y which passes through the eigenvalues of A such that $H = p(K)$. Since the nodes are symmetrically located with respect to the origin, the odd powers will cancel out leaving $\lceil \frac{n}{2} \rceil$ free real parameters. □

In Corollary 2 below we show how matrices of this type can arise in practice.

A converse of this theorem yields a way to generate real normal matrices with complex spectra. Choosing K to have a small bandwidth, the bandwidth of the arising normal matrix can be controlled with the degree of the polynomial used. However, forming polynomials in a matrix can be prohibitively expensive. It can also spoil the Toeplitz (or other) structure. Because the set of Toeplitz matrices is not closed under this operation, the FFT ideas can be used only by avoiding forming the polynomial explicitly.

More precisely, if $A \in \mathbb{C}^{n \times n}$ is a normal Toeplitz matrix, then after factoring the polynomial, matrix-vector products with $p(A)$ cost of order

$$\deg(p) O(n \log n) \tag{5}$$

operations. Since this is still impressive, let us consider a supporting matrix structure.

Forming polynomials in a Toeplitz matrix is actually a closed operation if Toeplitz matrices are viewed as persymmetric, i.e., symmetric with respect to the "anti-diagonal". Equivalently, if J is the permutation with ones on its anti-diagonal (the "backward identity" matrix [20]), then $A \in \mathbb{C}^{n \times n}$ is persymmetric if $JA^tJ = A$. Here A^t denotes the transpose of A.

Let $\mathcal{PS}$ denote the set of persymmetric matrices. This set has not received a lot of attention from the numerical linear algebra community. For some results, see [14]. For more pure linear algebraic considerations, see references in [1].

Proposition 2 *Let $A \in \mathbb{C}^{n \times n}$ be persymmetric. If p is a polynomial, then $p(A)$ is persymmetric.*

Proof. The matrix J is a unitary involution, that is, $J^{-1} = J = J^*$. Consequently, if $p(\lambda) = \alpha \prod_{j=1}^{k} (\lambda - \alpha_j)$ is the polynomial in its factored form, then we have

$$Jp(A)J = \alpha \prod_{j=1}^{k} J(A - \alpha_j I)J = p(A^t) = p(A)^t$$

and the claim follows. $\square$

Polynomials in a Toeplitz matrix thus remain persymmetric. In particular, if A is an invertible Toeplitz matrix, then its inverse is persymmetric. The set of persymmetric matrices is a subspace of $\mathbb{C}^{n \times n}$ but not a subalgebra although it is easy to verify that for two commuting persymmetric matrices the product is persymmetric. Moreover, one can also devise a nonsymmetric Lanczos iteration for persymmetric matrices; see, e.g., [12, 13].

Let $A^{at} = JA^tJ$ denote the anti-transpose of $A \in \mathbb{C}^{n \times n}$, i.e., the transpose with respect to the anti-diagonal. Clearly, normality is preserved in this operation. Then assign with the matrix a decomposition

$$A = \frac{A + A^{at}}{2} + \frac{A - A^{at}}{2} = P + S, \tag{6}$$

where P and S are the persymmetric and skew-persymmetric parts of A.

The persymmetric part P vanishes if and only if $JA^tJ = -A$. Then the eigenvalues of A are symmetrically located with respect to the origin (which is alarming for iteratively solving linear systems if A is not diagonally dominant). Also, then the anti-diagonal of A is zero and the even powers of A are again persymmetric while the odd powers remain skew-persymmetric.

The following is readily verified.

Proposition 3 *If $A = P + S \in \mathbb{C}^{n\times n}$ and $\lambda, \mu \in \mathbb{C}$, then $\lambda A + \mu I = \hat{P} + \hat{S}$ with $\hat{P} = \lambda S + \mu I$ and $\hat{S} = \lambda P$.*

Thus, the skew-persymmetric part of a matrix is translation invariant.

By a straightforward counting, the spaces of persymmetric and Hermitian persymmetric matrices in $\mathbb{C}^{n\times n}$ have real dimension $n(n+1)$ and $\frac{n(n+1)}{2}$, respectively.

Proposition 4 *The set of normal persymmetric matrices in $\mathbb{C}^{n\times n}$ is of real dimension $\frac{n(n+3)}{2}$.*

Proof. There are $\frac{n(n+1)}{2}$ real degrees of freedom to choose a Hermitian persymmetric matrix H. A polynomial p in H with real coefficients remains persymmetric. Thus, $H + ip(H)$ is a normal persymmetric matrix with, generically, $\frac{n(n+1)}{2} + n$ real parameters to choose. □

A naturally arising matrix nearness problem is solved with the splitting (6). By $||\cdot||$ we denote the spectral norm.

Theorem 2 *For $A \in \mathbb{C}^{n\times n}$ decomposed according to (6) there holds*

$$||A - P|| = \min_{X\in\mathcal{PS}} ||A - X|| .$$

Proof. By the unitary invariance of the spectral norm, for any $X \in \mathcal{PS}$ we have $||(X - A)^{at}|| = ||X - A||$. Then as in the proof of Fan and Hoffman [10], we get

$$||A - P|| = ||S|| = \frac{1}{2}\left|\left|A - X + (X^{at} - A^{at})\right|\right| \leq$$

$$\frac{1}{2}||A - X|| + \frac{1}{2}\left|\left|(X - A)^{at}\right|\right| = ||A - X||$$

and since $X \in \mathcal{PS}$ was arbitrary, the claim follows. □

Let $\mathcal{T}$ denote the set of Toeplitz matrices. The polynomial Toeplitz structure does not directly provide a means to approximate the skew-persymmetric part of a matrix since, for example, in the Frobenius norm the Pythagorean relation holds:

$$\min_{T\in\mathcal{T},\ p\in\mathcal{P}} ||A - p(T)||_{\mathcal{F}}^2 = \min_{T\in\mathcal{T},\ p\in\mathcal{P}} ||P - p(T)||_{\mathcal{F}}^2 + ||S||_{\mathcal{F}}^2 .$$

For strongly skew-persymmetric problems the FFT techniques can be used through "generalized" commutators. The proof is straightforward.

Proposition 5 *Let $M_0, M_1, \ldots, M_k \in \mathbb{C}^{n\times n}$ with M_0 persymmetric. Then $M_1^{at} M_2^{at} \cdots M_k^{at} - M_k M_{k-1} \cdots M_1$ and $M_1^{at} M_2^{at} \cdots M_k^{at} M_0 M_k M_{k-1} \cdots M_1$ are skew-persymmetric and persymmetric, respectively.*

In using the FFT ideas, this is of interest when each M_j is a low degree polynomial in a Toeplitz matrix. Note that with Hermitian M_j we have a skew-Hermitian and a Hermitian matrix, respectively.

The product in Proposition 5 had a symmetric structure. Without this symmetry we actually have a full generality with just two persymmetric matrices.

Theorem 3 *Any $A \in \mathbb{C}^{n\times n}$ is the product of two persymmetric matrices.*

Proof. We can employ the ideas from [34] after noticing that for any companion matrix C the product

$$CP = \begin{bmatrix} 0 & 0 & \cdots & \alpha_1 \\ 1 & 0 & \cdots & \alpha_2 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & \alpha_n \end{bmatrix} \begin{bmatrix} -1 & \alpha_n & \cdots & \alpha_2 \\ 0 & -1 & \cdots & \alpha_3 \\ \vdots & \ddots & \ddots & \alpha_n \\ 0 & \cdots & 0 & -1 \end{bmatrix} = B$$

is persymmetric. Thus, $C = BP^{-1}$, the product of two persymmetric matrices. Without loss of generality, assume A is similar to C (otherwise use the Frobenius canonical form and proceed analogously). Then

$$A = SCS^{-1} = SBP^{-1}S^{-1} = SBS^{at}S^{-at}P^{-1}S^{-1},$$

where both SBS^{at} and $S^{-at}P^{-1}S^{-1}$ are persymmetric by Proposition 5. $\square$

A natural problem is to characterize matrices that are representable as $A = p_1(T_1)p_1(T_2)$, for polynomials p_1 and p_2 and normal Toeplitz matrices T_1 and T_2. Then a linear system involving A can be solved by iteratively solving two consequtive linear systems involving normal matrices. In the Hermitian case this gives rise to centro-Hermitian matrices. A square matrix A is centro-Hermitian if $A^* = A^{at}$ [30, 18]. Then the decompositions (3) and (6) coincide.

Proposition 6 *If $A = p_1(H_1)\cdots p_k(H_k) \in \mathbb{C}^{n\times n}$, with polynomials p_j with real coefficients and with Hermitian Toeplitz matrices H_j, then A is centro-Hermitian.*

Proof. The product is a linear combination, with real coefficients, of terms of the form $H_1^{j_1}\cdots H_k^{j_k}$. It is enough to show that each of these terms is centro-Hermitian. But this follows from

$$\begin{aligned}\left(J(H_1^{j_1}\cdots H_k^{j_k})^t J\right)^* &= J(H_k^{j_k}\cdots H_1^{j_1})^t J \\ = J(H_1^{j_1})^t JJ\cdots JJ(H_k^{j_k})^t J &= H_1^{j_1}\cdots H_k^{j_k}.\end{aligned}$$

$\square$

By using the FFT techniques, matrix-vector products with these matrices cost of order $\sum_{j=1}^k \deg(p_j)O(n\log n)$ operations. Recall that a matrix A is the product of two Hermitian matrices if and only if A is similar to A^* [35].

The Hermitian case is of particular interest since Hermitian persymmetric matrices arise, e.g., in the integration of ODE's by multistep methods; see [8, Theorem 3.1].

Theorem 4 *If $A \in \mathbb{C}^{n\times n}$ is Hermitian, then so are P and S. Moreover, $S = \sum_{j=1}^{\frac{\mathrm{rank}(S)}{2}} F_j$, with Hermitian skew-persymmetric matrices F_j of rank 2 such that*

$$\min_{\mathrm{rank}(G)\le 2k} ||S-G|| = \left|\left|\sum_{j=k+1}^{\frac{\mathrm{rank}(S)}{2}} F_j\right|\right|,$$

for $k = 0,\ldots,\frac{\mathrm{rank}(S)}{2}$.

Proof. The first part of the claim is straightforward to verify. For the second part, since S is Hermitian and skew-persymmetric, its eigenvalues are symmetrically located on the real axis with respect to the origin. Let q_1 be an eigenvector of unit length related to the largest positive eigenvalue λ_1 of S. Then, because S is skew-persymmetric, $S^t J q_1 = -\lambda_1 J q_1$ holds, or equivalently, $S^* J\overline{q_1} = -\lambda_1 J\overline{q_1}$. Since S is Hermitian, $J\overline{q_1}$ is an eigenvector related to $-\lambda_1$. Being related to different eigenvalues, q_1 and $J\overline{q_1}$ are necessarily orthonormal. Therefore

$$F_1 = \lambda_1 q_1 q_1^* - \lambda_1 J\overline{q_1}(J\overline{q_1})^* = \lambda_1 q_1 q_1^* - \lambda_1 J\overline{q_1} q_1^t J \tag{7}$$

yields a best rank-2 approximation to S. Since $(q_1 q_1^*)^t = \overline{q_1} q_1^t$, the matrix F_1 is persymmetric as well. Continuing this construction with each positive-negative eigenvalue pair of S gives the claim. $\square$

To give a simple illustration of how the FFT ideas can be employed here, assume $A = P + S$ is a Hermitian matrix with a dominating persymmetric part. This could mean that either S has small norm, or $\operatorname{rank}(S) \ll n$. Then take T to be the nearest Hermitian Toeplitz matrix to P. If T is nonderogatory, it is straightforward to improve this with polynomials in T; first find the kernel of the linear mapping

$$X \to XT - TX \tag{8}$$

in $\mathbb{C}^{n\times n}$ and then compute its nearest element to P. In practice this approach is, of course, far too expensive. Instead, one should find

$$\min_{\deg(p)\ll n} ||Pb - p(T)b|| \tag{9}$$

for a vector $b \in \mathbb{C}^n$. This can be done with sparse matrix techniques; see [23]. The matrix $p(T)$ can be used, e.g., to precondition linear systems involving A. There is room for an improvement since the set of Hermitian matrices which are polynomials in a Hermitian Toeplitz matrix is of real dimension $3(n-1)$ (the reasoning: $2n-1$ real parameters to choose a Hermitian Toeplitz matrix T and $n-2$ for a polynomial with real coefficients; the constant and first order terms do not count).

Conversely, a dominating skew-persymmetric part is a sign of a more challenging problem.

Proposition 7 *If $A = P + S \in \mathbb{C}^{n\times n}$ is Hermitian such that $||P|| < ||S||$, then A is indefinite.*

Proof. Since S is skew-persymmetric and Hermitian, its eigenvalues are symmetrically located on the real axis with respect to the origin. Therefore, since both P and S are Hermitian and $||P|| < ||S||$ holds, A must have both negative and positive eigenvalues. $\square$

If we have additionally $\operatorname{rank}(P) \ll \operatorname{rank}(S)$, then A is even more seriously indefinite. Conversely, if S has small rank, then it may be worthwhile to further halve it as follows.

Theorem 5 *If $A = P + S \in \mathbb{C}^{n\times n}$ is Hermitian, then there exists a Hermitian matrix G of rank $\frac{\mathrm{rank}(S)}{2}$ such that $A - G$ is persymmetric.*

Proof. Choose

$$G = \sum_{j=1}^{\frac{\mathrm{rank}(S)}{2}} \lambda_j q_j q_j^* \tag{10}$$

from the construction initialized in (7). Then the skew-persymmetric part of the difference $A - G$ equals zero. $\square$

This yields a splitting $A = (A - G) + G$ of A, where the first part is Hermitian persymmetric and the second part is Hermitian. This decomposition is of interest because matrix-vector products with small rank matrices are inexpensive. More generally, consider those matrices A that can be represented as $A = M + G$, where $M = p(T)$ is a low degree polynomial in a Toeplitz matrix T and G is a small rank matrix. Performing matrix-vector products separately with the parts gives a bound

$$\deg(p)O(n\log n) + \mathrm{rank}(G)O(n) \tag{11}$$

on the complexity of matrix-vector products with A. A simple practical example of this can be given with a small rank perturbed (through its first $k \ll n$ columns) Toeplitz matrix

$$A = \begin{bmatrix} a_0 & \cdots & a_0 & 0 & 0 & \cdots & 0 \\ a_1 & \cdots & a_1 & a_0 & 0 & \cdots & 0 \\ a_2 & \cdots & a_2 & a_1 & a_0 & \cdots & 0 \\ \vdots & \ldots & \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix},$$

for which see [2].

While (10) was natural with the skew-persymmetric part of a Hermitian matrix, the following structure is with its persymmetric part.

Definition 1 *A vector $v \in \mathbb{C}^n$ is Hermitian if $\overline{Jv} = v$.*

This can be employed in generating small rank Hermitian persymmetric matrices. More precisely, if $v \in \mathbb{C}^n$ is Hermitian, then $J(vv^*)^t J = J\overline{v}(Jv)^t = vv^*$.

Proposition 8 *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian persymmetric matrix of rank 1. If $A \geq 0$, (resp. $A \leq 0$), then there exists a Hermitian vector $v \in \mathbb{C}^n$ such that $A = vv^*$ (resp. $A = -vv^*$).*

Proof. In the 2-by-2 case, if

$$vv^* = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}^* = \begin{bmatrix} |v_1|^2 & v_1\overline{v}_2 \\ v_2\overline{v}_1 & |v_2|^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ \overline{a}_{12} & a_{22} \end{bmatrix} = A,$$

then, due to persymmetry, $v_1 = |a_{12}|^{1/2} e^{i\alpha_1}$ and $v_2 = |a_{12}|^{1/2} e^{i\alpha_2}$, with $\arg(a_{12}) = \alpha_1 - \alpha_2$. Choosing $\alpha_2 = -\alpha_1 = \arg(a_{12})/2$ gives a Hermitian vector.

The 3-by-3 case follows by first using the 2-by-2 case with the corner entries of A to determine v_1 and v_3. Then using these with the equality of the $(1,2)$ and $(2,3)$ entries forces v_2 to be real.

The 4-by-4 case follows by first using the 2-by-2 case with the corner entries of A to determine v_1 and v_4. Then the equality of the $(1,3)$ and $(2,4)$ entries forces $v_2 = \overline{v}_3$. Continue this process by induction. $\square$

Note that with $w = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ and $v = \left[\begin{smallmatrix} i \\ -i \end{smallmatrix}\right]$ we have $ww^* = vv^*$ but only v is a Hermitian vector.

Theorem 6 *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian and persymmetric. Then A is unitarily diagonalizable as $A = \sum_{j=1}^{n} \lambda_j q_j q_j^*$ with each $q_j \in \mathbb{C}^n$ Hermitian.*

Proof. Consider a unitarily diagonalized $A = \sum_{j=1}^{n} \lambda_j q_j q_j^*$. If $Aq_j = \lambda_j q_j$, then

$$\overline{JA^tJq_j} = JA^*J\overline{q}_j = JAJ\overline{q}_j = \lambda_j \overline{q}_j.$$

Thus, $AJ\overline{q}_j = \lambda_j J\overline{q}_j$. If $J\overline{q}_j + q_j \neq 0$, then replace q_j with $w_j = J\overline{q}_j + q_j$ divided by its length. If $J\overline{q}_j + q_j = 0$, then replace q_j with $w_j = iq_j$ divided by its length. If all the eigenvalues of A are simple, this yields a unitarily diagonalized A with Hermitian eigenvectors.

If A has multiple eigenvalues, repeat the prescribed construction with those q_j for which q_j and $J\overline{q}_j$ are linearly dependent. For the remaining vectors q_j we have two dimensional subspaces $W_j = \operatorname{span}\{q_j, J\overline{q}_j\}$ which are spanned by the Hermitian vectors $w_j = J\overline{q}_j + q_j$ and $\tilde{w}_j = i(J\overline{q}_j - q_j)$. Among these vectors constructed, choose n linearly independent vectors and form a persymmetric Hermitian $E_k = \sum \epsilon_j^k w_j w_j^* + \sum \epsilon_l^k \tilde{w}_l \tilde{w}_l^*$. Choosing the real constants ϵ_j^k and ϵ_l^k appropriately, $A + E_k$ has simple eigenvalues such that $\lim_{k\to\infty} A + E_k = A$. Since each $A + E_k$ can be unitarily diagonalized with Hermitian eigenvectors, we have (after possibly passing to convergent subsequences) a unitarily diagonalized A with the properties claimed. $\square$

Aside from Hermitian problems, there are numerous applications where persymmetric matrices arise due to the fact that $A \in \mathbb{C}^{n\times n}$ is persymmetric if and only if JA is complex symmetric. For complex symmetric problems, see [11].

In this paper we do not consider algorithms for computing the decompositions proposed in this section. It is an interesting problem to devise an efficient algorithm in any of the nontrivial cases. It seems plausible that, at least in some instances, properties of the infinite dimensional problem being discretized can be employed. For example, with a Fredholm integral equation, (11) means, in the simplest case with $p(\lambda) = \lambda$, approximating the kernel with the sum of a convolution and a degenerate kernel. For some ideas in this direction, see [36]. The outcome of these attemps will eventually determine if these approaches are practical with the FFT ideas.

3 Two families of normal matrices

To have more options to execute optimal algorithms relying on a short term recurrence, we introduce two sparse matrix techniques for generating normal matrices. An aim is at having effortless ways to employ the FFT techniques together with normality.

3.1 Using embedding techniques

A Toeplitz matrix can be embedded in a circulant matrix of doubled size. If the circulant structure is relaxed, then any square matrix can, after dividing by its norm, be embedded in a unitary matrix; for a recent reference, see [6]. However, this is expensive since it involves computing a square root of a matrix. The following inexpensive 2-by-2 block structure generalizes the familiar case of $U = I$ for which see, e.g., [17, p.123] or [21, 1.6.11]. The proof is straightforward.

Proposition 9 *If $B \in \mathbb{C}^{n\times n}$, then $\left[\begin{smallmatrix} B & B^*U \\ U^*B^* & U^*BU \end{smallmatrix}\right] \in \mathbb{C}^{2n\times 2n}$ is normal for every unitary matrix $U \in \mathbb{C}^{n\times n}$.*

This is, however, not quite satisfactory since there is no flexibility regarding the spectrum (which is clearly independent on U). For example, $B = \left[\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ is invertible while its embedding is not. With the following we have more freedom with the eigenvalues.

Theorem 7 *Let $B \in \mathbb{C}^{n\times n}$ and assume U and V are mutually commuting unitary matrices commuting with B. Then*

$$M \equiv M_{U,V,\lambda}(B) = \begin{bmatrix} B & U(B-\lambda I)^* \\ V(B-\lambda I)^* & B \end{bmatrix} \in \mathbb{C}^{2n\times 2n} \tag{12}$$

is normal for any $\lambda \in \mathbb{C}$.

Proof. By the Putnam-Fuglede theorem, if B commutes with U, then B commutes with U^* as well; see, e.g., [17]. Consequently, B^* commutes with U and U^*. The same applies to the pair B and V. Then verifying that M is normal follows from a straightforward computation. $\square$

Aside from the complex scalar λ, this embedding is parametrized by two commuting unitary matrices commuting with the matrix B. The choices $U = e^{i\theta_1}I$ and $V = e^{i\theta_2}I$, with real θ_2 and θ_2, are obvious. If B is a polynomially normal matrix of low degree [26], then there are several alternatives to choose unitary matrices U and V such that the assumptions of the theorem are satisfied.

Corollary 1 *With $V = e^{i\alpha}U^*$ and $\alpha \in [0, 2\pi)$ the eigenvalues of $M - \lambda I$ are contained in the union of the lines*

$$y = -\frac{1+\cos(\alpha/2)}{\sin(\alpha/2)}x \text{ and } y = \frac{1-\cos(\alpha/2)}{\sin(\alpha/2)}x. \tag{13}$$

Proof. Without loss of generality, let $\lambda = 0$. Then we have

$$M_{U,V,0}(B)^2 = \begin{bmatrix} B^2 + UVB^{2*} & U(BB^* + B^*B) \\ V(B^*B + BB^*) & UVB^{2*} + B^2 \end{bmatrix}.$$

Therefore $M_{U,e^{i\alpha}U^*,0}(B)^2 - e^{i\alpha}M_{U,e^{i\alpha}U^*,0}(B)^{2*} = 0$ so that the function $z^2 - e^{i\alpha}\overline{z}^2$ annihilates the spectrum of $M_{U,e^{i\alpha}U^*,0}(B)$. The claim follows after solving for the corresponding equations of line; with $\alpha = 0$ we have lines $y = 0$ and $x = 0$. □

This particular choice yields a cross-like region centered at λ containing the eigenvalues of M. For iteratively solving a linear system involving matrices of this type, see [24, Example 3] where it is shown how the problem always separates into two alternating Hermitian Lanczos iterations. We find this quite remarkable since writing a complex linear system in its equivalent real form has a block structure that resembles M while giving rise to problematic eigenvalue configurations [11] with no hope of executing the separated Lanczos iterations mentioned.

The following sectoral containment relation for the field of values enforces a "semi" cross-like exclusion region for the eigenvalues. See also Figure 3.1.

Corollary 2 *Let the field of values of $B - \lambda I$ be located in a closed quadrant of $\mathbb{C}$ determined by the lines (13). Then, with $V = e^{i\alpha}U^*$ and $\alpha \in [0, 2\pi)$, the spectrum of $M - \lambda I$ is contained in the same quadrant on the lines (13).*

Proof. Without loss of generality we can assume that the field of values of B is in the right half-plane between the lines $y = \pm x$ and $\lambda = 0$. Due to Corollary 1, we only need to show that the eigenvalues of $M = M_{U,-U^*,0}$ are in the right half-plane. This we obtain by applying a unitary similarity to get

$$\mathrm{diag}(e^{i\pi/2}U^*,\ I)M\mathrm{diag}(e^{-i\pi/2}U,\ I) = \begin{bmatrix} B & e^{i\pi/2}B^* \\ e^{i\pi/2}B^* & B \end{bmatrix}. \qquad (14)$$

Thus, the eigenvalues of M are the union of the eigenvalues of $B \pm e^{i\pi/2}B^* = e^{i\pi/4}(e^{-i\pi/4}B \pm e^{i\pi/4}B^*)$. (Of course, if we knew these eigenvalues we would know the eigenvalues of M.) They are located in the right half-plane because of the assumption made on the location of the field of values of B. □

The embedding (12) yields a family of unitary matrices of doubled size once one unitary matrix is available (so that, when used recursively k times, we have unitary matrices of size $2^k n$ after starting with one of size n).

Corollary 3 *Assume $B \in \mathbb{C}^{n\times n}$ is unitary, $V = -U^*B^4$ and $\lambda = 0$. Then $\frac{1}{\sqrt{2}}M$ is unitary whose spectrum independent on U.*

Proof. By demanding $M_{U,V,0}M^*_{U,V,0} = I$, we obtain from the $(2,1)$-block the condition $VB^{2*} + U^*B^2 = 0$ which is equivalent to $V = -U^*B^4$. Then the $(1,1)$-block yields the factor $\frac{1}{\sqrt{2}}$. Regarding the second claim, with this choice we have

$$M_{U,-U^*B^4,0} = \begin{bmatrix} B & UB^* \\ -U^*B^3 & B \end{bmatrix} = \begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} B & B^* \\ -B^3 & B \end{bmatrix}\begin{bmatrix} U^* & 0 \\ 0 & I \end{bmatrix}$$

which proves the invariance of the spectrum. □

Since the set of commuting unitary matrices forms an Abelian group, natural alternatives for U are powers of B and B^*. Then the number of choices equals the order of the arising group.

Another natural family of unitary matrices is obtained with a Hermitian matrix H. Then form B and U by performing Cayley transformations of polynomials with real coefficients in H.

Let $\mathcal{P}_o$ denote the set of polynomials with odd powers. Proposition 5 is of use with the following.

Proposition 10 *Let $B \in \mathbb{C}^{n\times n}$ be skew-persymmetric and U and V persymmetric both. Then, with $\lambda = 0$ and $p \in \mathcal{P}_o$, the matrix $p(M)$ is skew-persymmetric.*

Proof. Since B is skew-persymmetric, the $(1,1)$-block is the anti-reflection of the $(2,2)$-block across the anti-diagonal of M. Moreover, if B is skew-persymmetric, then so is B^*. Since U and B^* commute,

$$J(UB^*)^tJ = J(B^*U)^tJ = JU^t(B^*)^tJ = JU^tJJ(B^*)^tJ = -UB^*,$$

so that UB^* is skew-persymmetric. The same reasoning applies to VB^* which establishes the claim since odd powers of skew-persymmetric matrices remain skew-persymmetric. □

To generate normal persymmetric matrices we have the following option with any Toeplitz matrix.

Proposition 11 *Let $B \in \mathbb{C}^{n\times n}$ be persymmetric and nonderogatory. Then $p(M)$ is persymmetric for any polynomial p.*

Proof. Since B is nonderogatory, U and V are necessarily polynomials in B to commute with B. Therefore both U and V are persymmetric. Moreover, since B is persymmetric, then so is $(B-\lambda I)^*$. Because U and V commute with B^*, both $U(B-\lambda I)^*$ and $V(B-\lambda I)^*$ are persymmetric and the claim follows by using Proposition 2. □

For any $B \in \mathbb{C}^{n\times n}$ and a polynomial p the $(1,1)$-block and $(2,2)$-block of $p(M)$ equal. Because of this block-persymmetry, a necessary condition for having reasonable normal approximations with $p(M)$ to a given matrix $A = \left[\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}\right] \in \mathbb{C}^{2n\times 2n}$ is that the difference $A_{11} - A_{22}$ is small (or it has fast decaying singular values). Then combining normality with the FFT ideas amounts to choosing a Toeplitz matrix B and a translation parameter λ, with $U = e^{i\theta_1}$, $V = e^{i\theta_2}$, according to

$$\min_{B\in\mathcal{T},\ \lambda\in\mathbb{C},\ \theta_1,\theta_2\in\mathbb{R}} \Big(||A_{11} - B||_{\mathcal{F}}^2 + \big|\big|A_{12} - e^{i\theta_1}(B-\lambda I)^*\big|\big|_{\mathcal{F}}^2 + \\ + \big|\big|A_{21} - e^{i\theta_2}(B-\lambda I)^*\big|\big|_{\mathcal{F}}^2 + ||A_{22} - B||_{\mathcal{F}}^2 \Big) . \tag{15}$$

This is straightforward to find. This approximation can be polynomially improved by using the criterion (9) with sparse matrix techniques.

The converse of the fact that any matrix has a normal embedding has a quite surprising consequence. Namely, a normal matrix can have an arbitrarily nonnormal compression, i.e., if $Q \in \mathbb{C}^{2n\times n}$ has orthonormal columns corresponding to the $(1,1)$-block of M, then Q^*MQ is unitarily similar to B. This is in strong contrast with the Hermitian case since every compression of a Hermitian matrix remains Hermitian. For an iterative method producing approximations preserving normality for normal matrices, see [22]. For almost normal compressions, see [27].

On the positive side, the fact that a normal matrix can have a nonnormal compression can be very useful. To see this, consider solving a linear system $Ax = b$, with $A \in \mathbb{C}^{n\times n}$ and for $b \in \mathbb{C}^n$. Associate with the problem an invertible embedding with

$$M^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \in \mathbb{C}^{2n\times 2n}. \tag{16}$$

Now the matrix C_{11}, which can be arbitrary nonnormal, can be taken as a preconditioner for the linear system. The trick is that even though C_{11} is not assumed to be explicitly available, matrix-vector products with it can be computed through solving linear systems involving M. This, in turn, can be done efficiently by using iterative methods for normal matrices. Then the problem arises, how to choose an embedding to this end. The following elementary remark is of use.

Proposition 12 *Under the assumptions of Theorem 7,*

$$C_{11}^{-1} = B - UV(B-\lambda I)^* B^{-1}(B-\lambda I)^* \tag{17}$$

as long as the inverses exist.

Proof. For having the inverse (16), consider the defining equation

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} B & U(B-\lambda I)^* \\ V(B-\lambda I)^* & B \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \tag{18}$$

giving the conditions $C_{11}B + C_{12}V(B-\lambda I)^* = I$ and $C_{11}U(B-\lambda I)^* + C_{12}B = 0$ from the first block-row. Solving for C_{12} in terms of C_{11} from the second equation gives $C_{12} = -C_{11}U(B-\lambda I)^*B^{-1}$. Inserting this to the first equation then yields the inverse of C_{11} as claimed. □

Thus, in principle, any linear system can be solved by solving a linear system whose coefficient matrix is a normal embedding. In practice the matrix B along with U, V and λ should be chosen such that (17) approximates the original matrix $A \in \mathbb{C}^{n\times n}$ well. An equality seems to be difficult to attain inexpensively.

To give an example on how to combine here the FFT ideas with normality, assume A is a nonnormal Toeplitz matrix with the symbol g. Fix $U = e^{i\theta_1}$ and $V = e^{i\theta_2}$, for $\theta_1, \theta_2 \in \mathbb{R}$, and $\lambda \in \mathbb{C}$ appropriately. Then, in view of the relation (17), a natural approximation problem amounts to finding f such that

$$\frac{f(z)^2 - e^{i\theta_1}e^{i\theta_2}(\overline{f(z)} - \overline{\lambda})^2}{f(z)} = g(z) \tag{19}$$

holds for z in the unit circle. This is a second degree equation for (the real and imaginary parts of) f and thereby readily solvable. The matrix B is then chosen to be the Toeplitz matrix corresponding to the symbol f.

For using the symbol in a more standard fashion through $1/f$ in Toeplitz preconditioning, see [4, 31]. A drawback of this approach is that the construction of the preconditioner can be expensive due to the fact that the Fourier coefficients of $1/f$ are not readily available. There is more flexibility with (19).

Example 1. Assume $g(z) = g_1(z) + ig_2(z)$ attains values in a sector of angle $\pi/2$ not containing the origin, say, between the lines $y = x$ and $y = -x$ in the right half plane. Then with $\lambda = 0$, $\theta_1 = 0$ and $\theta_2 = \pi$ the solution to (19) reads

$$f(z) = \frac{|g(z)|^2}{g_1(z)^2 - g_2(z)^2}\frac{\overline{g(z)}}{2}. \tag{20}$$

If values of g are close to the lines $y = \pm x$, then choose a negative λ and recompute f. Like with $1/f$, the Toeplitz matrix with this symbol can be expensive to construct. However, there are approximations with varying accuracy. The coarsest one corresponds to replacing the first factor, which attains only real values, with a constant $c \in \mathbb{R}$. With $c = 1$ we have $\overline{g}/2$ which gives simply $B = A^*/2$. To improve this, replace some of the diagonals of B with the exact Fourier coefficients of f.

For $\lambda = 0$, $\theta_1 = 0$, and $\theta_2 = \pi$, the choice $B = A^*/2$ in the previous example is reasonable more generally if $A = dH + E$ with $d \in \mathbb{C}$, a Hermitian H and $E \in \mathbb{C}^{n\times n}$ of moderate size. To see this, consider just the leading term $(dH/2)^*$. Inserting it in (17) gives $\frac{d^2+\overline{d}^2}{\overline{d}}H$, which is ideal in view of iterative methods, as long as the constant is nonzero.

Example 2. Assigning numerical values to matrices of Example 1, let $A = A_1 + A_2 \in \mathbb{C}^{500\times 500}$, where A_1 is a Hermitian Toeplitz matrix and A_2 is a random Toeplitz matrix scaled to have norm 20. Rounding to four digits, the smallest and the largest eigenvalues of A_1 are 13.76 and 162.6, respectively. See Figure 3.1. We set $B = A^*/2$. Then the condition numbers of A and $M_{I,-I,0}$ are 10.75 and 14.61, respectively (so that the condition number of A^*A is 115.5). This gives us $||C_{11}A - I|| = 0.2742$ which is quite impressive. The computations were performed with `Matlab` [32].

Although the 3-term recurrence of [24] is almost ideal for solving these linear systems, in section 4 we introduce another very natural optimal method for normal matrices of this section. Since the eigenvalues of the embeddings considered are located on a second degree algebraic curve, we have a 5-term recurrence.

Here we have another opportunity to combine normality with the FFT ideas by the fact that (17) is not a Toeplitz matrix for a Toeplitz matrix B.

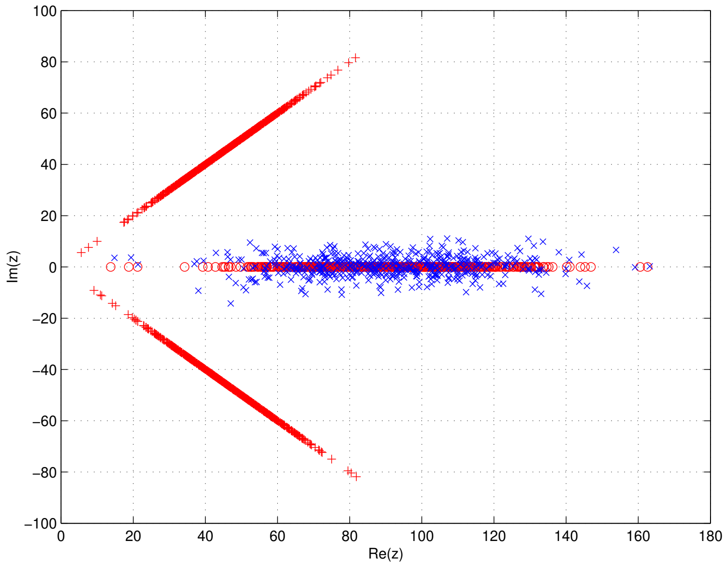

Figure 1: From Example 2 the eigenvalues of matrices A_1, A and $M_{I,-I,0}$ depicted by $'o'$, $'x'$ and $'+'$, respectively.

Thus, consider finding a Toeplitz matrix B such that (17) approximates a given matrix $A \in \mathbb{C}^{n \times n}$. The arising structure is again persymmetry.

Proposition 13 *Under the assumptions of Proposition 12, if U, V and B are additionally persymmetric, then (17) is persymmetric.*

Proof. Since $JC_{11}^{-t}J$ equals

$$J\big(B^t - U^t JJV^t JJ((B-\lambda I)^*)^t JJB^{-t}JJ((B-\lambda I)^*)^t\big)J, \tag{21}$$

the claim follows from persymmetry of U, V and B, and by the fact that persymmetry is preserved in inversion and in taking the adjoint. $\square$

Since persymmetry is again preserved, in view of Toeplitz preconditioning, assume $A = p(T)$ (or approximatively) for a Toeplitz matrix $T \in \mathbb{C}^{n \times n}$ and for a low degree polynomial p. If g is the symbol of T, then it seems natural to replace the right-hand side of (19) with the symbol $p(g)$ and solve for f.

In (21) we used the fact that the mapping

$$P \to M^{at}PM \tag{22}$$

preserves persymmetry for any fixed $M \in \mathbb{C}^{n \times n}$. An interesting problem is to characterize those persymmetric matrices $P \in \mathbb{C}^{n \times n}$ for which $M^{at}PM$ is a Toeplitz matrix for some simple invertible M like for a Toeplitz or a diagonal matrix. This resembles scaling with the aim now at having a Toeplitz matrix. A straightforward approach in the diagonal case is as follows.

Example 3. Let $A \in \mathbb{C}^{n \times n}$ and assume that for $j = 1, 2$ the diagonal matrices $D_j = \operatorname{diag}(d_1^j, \ldots, d_n^j)$ satisfy $|D_j| \geq cI$ for a fixed $c > 0$. The (j,k)-entry of D_1AD_2 is $d_j^1 a_{jk} d_k^2$ so that a nearest Toeplitz matrix T to D_1AD_2 is readily computed. In the Frobenius norm we just take the average of every diagonal. Then

$$\min_{|D_j|\geq cI} \operatorname{dist}\big(D_1 A D_2, \mathcal{T}\big) \tag{23}$$

gives a criterion for choosing the diagonal matrices. See [36] for an example of this when A is a "generalized" Hilbert matrix with zero diagonal and $a_{jk} = (z_j - z_k)^{-1}$ otherwise for a given set $\{z_l\}_{l=0}^{n-1} \subset \mathbb{C}$.

3.2 Using the Kronecker product

The Kronecker product can be employed to have more variation in block structures aside from forming polynomials in matrices (12). To this end, recall that the Kronecker product $N_1 \otimes N_2$ is normal if the factors $N_1 \in \mathbb{C}^{k\times k}$ and $N_2 \in \mathbb{C}^{l\times l}$ are [21]. It is straightforward to verify that then

$$p_\otimes(N_1, N_2) = \sum q_j(N_1) \otimes p_j(N_2) \tag{24}$$

is normal, where q_j and p_j are polynomials of degree $k-1$ and $l-1$ at most, respectively.

Proposition 14 *Let $N_1 \in \mathbb{C}^{k\times k}$ and $N_2 \in \mathbb{C}^{l\times l}$ be normal. Then $p_\otimes(N_1, N_2)$ is normal.*

Proof. Mimic the proof of [21, Theorem 4.2.12]. □

Incorporating this with the methods introduced earlier we have further alternatives to generate normal matrices by forming $p_\otimes(N_1, N_2)$ with normal matrices N_1 and N_2 generated so far. Obviously, there are $\deg(N_1)\deg(N_2)$ complex parameters to choose in forming $p_\otimes(N_1, N_2)$.

Giving $p_\otimes(N_1, N_2)$ in terms of polynomials in N_1 and N_2 is not standard as opposed to the representation (2). However, it is better suited to numerical computations. First, power bases are not numerically stable to generate. Second, if we have $k \ll l$ with $kl = n$, then it is feasible to compute polynomials in N_1 explicitly. In this manner matrix-vector multiplications with $p_\otimes(N_1, N_2)$ can be computed without forming polynomials in N_2 explicitly. Consequenly, if N_2 is a polynomial in a normal Toeplitz matrix, the FFT techniques become available.

The Kronecker product of two persymmetric matrices is readily seen to be persymmetric and therefore so is $p_\otimes(N_1, N_2)$. Similarly, we have the following.

Proposition 15 *Assume $N_1 \in \mathbb{C}^{k\times k}$ and $N_2 \in \mathbb{C}^{l\times l}$ are skew-persymmetric and persymmetric, respectively. Let $q_j = q_j^e + q_j^o$, where q_j^e and q_j^o are polynomials with even and odd powers, respectively. Then*

$$P = \sum q_j^e(N_1) \otimes p_j(N_2) \quad \textit{and} \quad S = \sum q_j^o(N_1) \otimes p_j(N_2),$$

for $p_\otimes(N_1, N_2)$.

The structure introduced is natural while dealing with linear systems of the form

$$A_1XB_1 + \cdots + A_pXB_p = C, \tag{25}$$

with matrices $A_j \in \mathbb{C}^{k\times k}$ and $B_j \in \mathbb{C}^{l\times l}$, for $j = 1,\ldots,p$. The Sylvester equation, of which the Lyapunov equation is a special case, belongs to this class of linear systems with $p = 2$ and $B_1 = A_2 = I$. Another example is the Stein equation with $p = 2$ and $B_1 = A_1^*$ and $A_2 = -B_2 = I$. For problems where sums of Kronecker products with 3 and 4 terms arise, see [3].

The linear system (25) can be written in the standard form with a sum of Kronecker products; see, e.g., [21, Section 4.3]. The construction of a normal preconditioner $p_\otimes(N_1, N_2)$ to approximate the corresponding coefficient matrix amounts to approximating B_j^t and A_j, for $j = 1,\ldots,p$, with polynomials in N_1 and N_2, respectively. It is an interesting problem how to choose N_1 and N_2 to this end.

4 An optimal iterative method for normal matrices

In this final section we introduce an iterative method for the normal embedding proposed in section 3. For this purpose we need an appropriate function class. So far we have only considered polynomials which is not sufficent. With normal matrices it is more natural to employ polyanalytic polynomials because then it is possible to use real analytic computational techniques [25, 27].

Definition 2 *Polyanalytic polynomials are functions of the form*

$$p(z) = \sum_{0\le j+l\le k} c_{j,l}z^l\overline{z}^j, \tag{26}$$

with $c_{j,l} \in \mathbb{C}$.

Polyanalytic polynomials of the form $z^j\overline{z}^l$ are called polyanalytic monomials and an order $>$ among them is set as follows. Let $z^{j_1}\overline{z}^{l_1}$ and $z^{j_2}\overline{z}^{l_2}$ be two polyanalytic monomials. If $j_1 + l_1 > j_2 + l_2$, then $z^{j_1}\overline{z}^{l_1} > z^{j_2}\overline{z}^{l_2}$. If $j_1 + l_1 = j_2 + l_2$ and $j_1 > j_2$, then $z^{j_1}\overline{z}^{l_1} > z^{j_2}\overline{z}^{l_2}$. With an order among the polyanalytic monomials we define the minimal polyanalytic polynomial $p_{j,l}$ of a normal $N \in \mathbb{C}^{n\times n}$ to be the monic polyanalytic polynomial of least degree $j + l$ annihilating N. For example, the matrices of Corollary 1 had the minimal polyanalytic polynomial $p_{2,0}(z) = (z-\lambda)^2 - e^{i\alpha}(\overline{z}-\overline{\lambda})^2$, where the sub-indeces refer to its leading term z^2.

The minimal polyanalytic polynomial can be computed by generating an orthonormal basis of the Krylov subspace

$$\mathcal{K}(N; b) = \operatorname{span}\{b, N^*b, Nb, N^{*2}b, NN^*b, N^2b, \ldots\}, \tag{27}$$

with a vector $b \in \mathbb{C}^n$. Note the unusual ordering of the matrix-vector products. A numerically stable implementation of this process has been carefully described in [27].

If the degree of the minimal polyanalytic polynomial $p_{j,l}$ is moderate, then linear systems involving N can be iteratively solved with an optimal short term recurrence. The classical Hermitian Lanczos method is a particular case of this with its optimal 3-term recurrence. So is the method of Jagels and Reichel [29]. Since the idea can be readily generalized, we consider the case of $p_{2,0}$ only. Then, for any $j, k \geq 0$, the vectors $N^{2+k}N^{*j}\hat{q}_0$ are linearly dependent on the vectors to their left in (27). By repeating the steps used in [24], it is straightforward to devise a 5-term recurrence which minimizes the residual over the Krylov subspace (of arbitrary high dimension) generated so far. The corresponding optimality condition is bounded by an approximation problem on the spectrum $\sigma(N)$ of N according to

$$\min_{\deg(p)\leq k} ||Np(N)b - b|| \leq \min_{\deg(p)\leq k} \max_{\lambda\in\sigma(N)} |\lambda p(\lambda) - 1| \, ||b|| \,, \tag{28}$$

where p belongs to the set of polyanalytic polynomials. Recall that with GMRES p varies only among polynomials.

An implementation of the method is as follows.

ALGORITHM 1. "For solving $Nx = b$ for a normal $N \in \mathbb{C}^{n\times n}$ with $p_{2,0}$."

$q_0 = b/\,||N\,b||$

$q_1 = N^*\,q_0 - (N^*N\,q_0, N\,q_0)q_0, \quad q_1 = q_1/\,||Nq_1||$

$q_2 = N\,q_0 - (N^2\,q_0, N\,q_1)q_1 - (N^2\,q_0, N\,q_0)q_0, \quad q_2 = q_2/\,||N\,q_2||$

$x = (b, N\,q_2)q_2 + (b, N\,q_1)q_1 + (b, N\,q_0), \quad r = b - N\,x$

for $j = 1 : K$

  $q = N^*\,q_{2j-1}$

  for $s = 0 : 3$

    $\alpha = (N^*N\,q_{2j-1}, N\,q_{2j-s}), \quad q = q - \alpha q_{2j-s}$

  end

  $q_{2j+1} = q/\,||N\,q||$

  $\alpha = (r, N\,q_{2j+1})q_{2j+1}$

  $x = x + \alpha q_{2j+1}$

  $r = r - \alpha N\,q_{2j+1}$

  $q = N\,q_{2j-1}$

  for $s = 0 : 3$

    $\alpha = (N^2\,q_{2j-1}, N\,q_{2j+1-s}), \quad q = q - \alpha q_{2j+1-s}$

  end

  $q_{2j+2} = q/\,||N\,q||$

  $\alpha = (r, N\,q_{2j+2})q_{2j+2}$

  $x = x + \alpha q_{2j+2}$

  $r = r - \alpha N\,q_{2j+2}$

end

Note that inside the for-loop we always perform two iteration steps.

Since p varies among the set of polyanalytic polynomials, Algorithm 1 is guaranteed to yield simultaneously improved GMRES and the normal equations approximations in the following sense.

Proposition 16 *Let $N \in \mathbb{C}^{n\times n}$ be normal with the minimal polyanalytic polynomial $p_{2,0}$. Then the norm of the residual at the k^{th} step with Algorithm 1 does not exceed the norm of the GMRES or the normal equations residual after $k/2$ and $(k-2)/4$ steps, respectively.*

Proof. For any $s, t \geq 0$ the vectors $N^{2+s}N^{*t}\hat{q}_0$ are linearly dependent on the vectors to their left in the subspace (27). Therefore, by disregarding the redundant matrix-vector multiplications, at the step $k = 2j$ the vectors $\{N^l b\}_{l=0}^{j}$ belong to the subspace over which Algorithm 1 minimizes the residual. These vectors constitute the GMRES approximation. Similarly for the normal equations, at the step $k = 4j + 2$ the vectors $\{(NN^*)^l b\}_{l=0}^{j}$ belong to the subspace. □

This also implies that the extreme failures of GMRES versus CGN, and vice versa, described in [33] do not take place with Algorithm 1.

5 Conclusions

Various matrix structures supporting both normality and using the FFT techniques have been studied. Polynomials in Toeplitz matrices remain persymmetric and therefore, regarding iterative methods, persymmetric matrices yield a natural framework for considering extensions of the FFT ideas. Methods for generating normal matrices were introduced. The one based on embedding matrices in normal matrices allows us to solve nonnormal linear systems via systems involving normal matrices. Then the problem how to choose an embedding matrix arises. With Toeplitz related matrices persymmetric structure reappears so that we can employ symbols to this end. Finally, an optimal 5-term recurrence was introduced for solving linear systems involving normal matrices arising in this context.

Efficient computability of these matrix structures provides challenging and interesting problems. For practical purposes this will eventually determine whether these approaches can be useful.

References

[1] A. Andrew, *Centrosymmetric matrices,* SIAM Rev., 40, (1998), pp. 697–698.

[2] D. Bini and B. Meini, *Exploiting the Toeplitz structure in certain queueing problems,* Calcolo, 33 (1996), pp. 289–305.

[3] R.N. Chan and W.K. Ching *Circulant preconditioners for stochastic automata networks,* Numer. Math. 87 (2000), pp. 35–57.

[4] R.N. Chan and K-P. Ng, *Toeplitz preconditioners for Hermitian Toeplitz systems,* Lin. Alg. Appl., 190 (1993), pp. 181–208.

[5] R.N. Chan and M.K. Ng, *Conjugate gradient methods for Toeplitz systems*, SIAM Rev. 38 (1996), pp. 427–482.

[6] M-D. Choi and C-K. Li, *Constrained unitary dilations and numerical ranges,* J. Oper. Th., 46 (2001), pp. 435-447.

[7] P. Davis, Circulant matrices, *John Wiley and Sons, New York, 1979.*

[8] T. Eirola and J.M. Sanz-Serna, *Conservation of integrals and symplectic structure in the integration of differential equations by multistep methods,* Numer. Math., 61 (1992), pp. 281–290.

[9] L. Elsner and Kh.D. Ikramov, *On a condensed form for normal matrices under finite sequence of elementary similarities,* Lin. Alg. Appl., 254 (1997), pp. 79–98.

[10] K. Fan and A. Hoffman, *Some metric inequalities in the space of matrices,* Proc. Amer. Math. Soc. 6, (1955), pp. 111–116.

[11] R. Freund, *On conjugate-gradient type methods for linear systems with complex-symmetric coefficient matrices,* SIAM J. Sci. Stat. Comput., 13 (1992), pp. 425-448.

[12] R. Freund, *Lanczos-type algorithms for structured non-Hermitian eigenvalues problems,* Proceedings of the Cornelius Lanczos International Centenary Conference, (Raleigh, NC, 1993), pp. 243-245, SIAM, Philadelphia, PA, 1994.

[13] R. Freund, G Golub and N. Nachtigal *Iterative solution of linear systems,* Acta Numerica, 1, pp. 57-100, 1992.

[14] M. Goldstein, *Reduction of the eigenproblem for Hermitian persymmetric matrices,* Math. Comp., 28 (1974), pp. 237–238.

[15] A. Greenbaum, Iterative Methods for Solving Linear Systems, *Frontiers in Applied Mathematics, SIAM, Philadelphia,* 1997.

[16] R. Grone, C.R. Johnson, E.M. Sa and H. Wolkowicz, *Normal matrices,* Lin. Alg. Appl., 87 (1987), pp. 213–225.

[17] P. Halmos, A Hilbert space problem book, *2nd Ed. Grad. Texts in Math. 19, Springer-Verlag, New York-Berlin,* 1982.

[18] R. Hill, R. Bates and S. Waters, *On centro-Hermitian matrices,* SIAM J. Matrix Anal. Appl. 11 (1990), pp. 128–133.

[19] Y.P. Hong and R.A. Horn, *The Jordan canonical form of a product of a Hermitian and a positive semidefinite matrix,* Lin. Alg. Appl., 147 (1991), pp. 373-386.

[20] R.A. Horn and C.R. Johnson, Matrix Analysis, *Cambridge Univ. Press, Cambridge,* 1987.

[21] R.A. Horn and C.R. Johnson, Topics in Matrix Analysis, *Cambridge Univ. Press, Cambridge,* 1991.

[22] M. Huhtanen, *A stratification of the set of normal matrices,* SIAM J. Matrix Anal. Appl., 23 (2001), pp. 349–367.

[23] M. Huhtanen, *A matrix nearness problem related to iterative methods*, SIAM J. Numer. Anal., 39 (2001), pp. 407–422.

[24] M. Huhtanen, *A Hermitian Lanczos method for normal matrices*, SIAM J. Matrix Anal. Appl., 23 (2002), pp. 1092-1108.

[25] M. Huhtanen, *Orthogonal polyanalytic polynomials and normal matrices,* Math. Comp., to appear.

[26] M. Huhtanen, *Aspects of nonnormality related to itrative methods,* manuscript, 2002.

[27] M. Huhtanen and R.M. Larsen, *Exclusion and inclusion regions for the eigenvalues of a normal matrix*, SIAM J. Matrix Anal. Appl., 23 (2002), pp. 1070-1091.

[28] KH.D. Ikramov, *Comments on: "Normal Toeplitz matrices" [SIAM J. Matrix Anal. Appl., 17 (1996), pp. 1037–1043] by D. R. Farenick, M. Krupnik, N. Krupnik, and W. Y. Lee.,* SIAM J. Matrix Anal. Appl. 18 (1997), p. 518.

[29] C. Jagels and L. Reichel, *A fast minimal residual algorithm for shifted unitary matrices,* Numer. Linear Algebra Appl., 1 (1994), pp. 555–570.

[30] A. Lee, *Centro-Hermitian and skew-centro-Hermitian matrices,* Lin. Alg. Appl., 29 (1980), pp. 205–210.

[31] J. Malinen, *Properties of iteration of Toeplitz operators with Toeplitz preconditioners,* BIT, 38 (1998), pp. 356–371.

[32] Mathworks, *Matlab*, `www.mathworks.com/products/matlab`.

[33] N.M. Nachtigal, S. Reddy and L.N. Trefethen, *"How fast are nonsymmetric matrix iterations?"*, SIAM J. Matrix Anal. Appl., 3 (1992), pp. 778–795.

[34] H. Radjavi, *Products of Hermitian matrices and symmetries,* Proc. A.M.S. 21 (1969), pp. 369-372.

[35] H. Radjavi and J.P. Williams, *Products of self-adjoint operators,* Michigan Math. J., 16 (1969), pp. 177-185.

[36] L. Reichel, *A matrix problem with application to rapid solution of integral equations,* SIAM J. Sci. Statist. Comput., 11 (1990), pp. 263–280.

[37] G. Strang, *A proposal for Toeplitz matrix calculations*, Studies in Appl. Math., 74 (1986), pp. 171-176.

(continued from the back cover)

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS
RESEARCH REPORTS

The list of reports is continued inside. Electronical versions of the reports are available at *http://www.math.hut.fi/reports/* .

A461 Tuomas Hytönen
Vector-valued wavelets and the Hardy space $H^1(R^n; X)$
April 2003

A460 Jan von Pfaler , Timo Eirola
Numerical Taylor expansions for invariant manifolds
April 2003

A459 Timo Salin
The quenching problem for the N-dimensional ball
April 2003

A458 Tuomas Hytönen
Translation-invariant Operators on Spaces of Vector-valued Functions
April 2003

A457 Timo Salin
On a Refined Asymptotic Analysis for the Quenching Problem
March 2003

ISBN 951-22-6152-9
ISSN 0784-3143
Inst. of Math. HUT, Espoo, 2002