FACTORING MATRICES INTO THE PRODUCT OF CIRCULANT AND DIAGONAL MATRICES

MARKO HUHTANEN* AND ALLAN PERÄMÄKI†

Abstract. A generic matrix $A \in \mathbb{C}^{n\times n}$ is shown to be the product of circulant and diagonal matrices with the number of factors being $2n-1$ at most. The demonstration is constructive, relying on first factoring matrix subspaces equivalent to polynomials in a permutation matrix over diagonal matrices into linear factors. For the linear factors, the sum of two PD matrices is factored into the product of two diagonal matrices and a circulant matrix. Extending the monomial group, low degree polynomials in a permutation matrix over diagonal matrices and their permutation equivalences constitutes a fundamental sparse matrix structure. Matrix analysis gets largely done in terms of permutations only.

Key words. circulant matrix, diagonal matrix, monomial group, sum of PD matrices, polynomial factoring, permutation matrix analysis, sparsity, polynomial permutation degree

AMS subject classifications. 15A23, 12D05, 05C50

1. Introduction.

There exists an elegant result, motivated by applications in optical image processing, stating that any matrix $A \in \mathbb{C}^{n\times n}$ is the product of circulant and diagonal matrices [14, 16].[1] In this paper it is shown that, generically, $2n-1$ factors suffice. (For various aspects of matrix factoring, see [12].) The demonstration is constructive, relying on first factoring matrix subspaces equivalent to polynomials in a permutation matrix over diagonal matrices into linear factors. Located on the borderline between commutative and noncommutative algebra, such subspaces are shown to constitute a fundamental sparse matrix structure of polynomial type. Then for the linear factors, a factorization result for the sum of two PD matrices is derived.

A scaled permutation, also called a PD matrix, is the product of a permutation and a diagonal matrix. In the invertible case we are dealing with the monomial group, giving rise to the sparsest possible nonsingular matrix structure. A way to generalize this is to allow more nonzero entries per line by considering sums of PD matrices. The sum of two PD matrices can be analyzed in terms of permutation equivalence which turns out to be instrumental for extending the structure. Although the notion of permutation equivalence is graph theoretically nonstandard, combinatorial linear algebraically it is perfectly natural [2, p. 4]. There arises a natural concept of cycles which can be used to show that the inverse of a nonsingular sum of two PD matrices carries a very special structure and can be inexpensively computed.

To extend the set of sums of two PD matrices in a way which admits factoring, a polynomial structure in permutations is suggested. That is, let P be a permutation matrix and denote by p a polynomial over diagonal matrices. Define matrix subspaces of $\mathbb{C}^{n\times n}$ as

$$P_1 \left\{ p(P) \,\middle|\, \deg(p) \le j \right\} P_2 \tag{1.1}$$

with fixed permutations P_1 and P_2. This provides a natural extension by the fact that the case $j = 0$ corresponds to PD matrices while $j = 1$ yields the sums of

* Division of Mathematics, Department of Electrical and Information Engineering, University of Oulu, 90570 Oulu 57, Finland, (Marko.Huhtanen@aalto.fi).

† Department of Mathematics and Systems Analysis, Aalto University, Box 1100 FIN-02015, Finland, (allan.peramaki@aalto.fi).

[1] In particular, any unitary matrix $A \in \mathbb{C}^{n\times n}$ can be interpreted as being a diffractive optical system.

two PD matrices. The case $j = 2$ covers, e.g., finite difference matrices, including periodic problems. In this manner, whenever $j \ll n$, the sparsity pattern of such a matrix subspace carries an intrinsic polynomial structure which can be used to analyze sparsity more generally in terms of the so-called polynomial permutation degree. In particular, now matrix analysis gets largely done in terms of permutations. Namely, completely analogously to univariate complex polynomials, these subspaces admit factoring. To factor (1.1) into linear factors, it turns out that it suffices to consider the problem of factoring polynomials in the cyclic shift[2] over diagonal matrices.

Let P thus be the cyclic shift and set $P_1 = P_2 = I$. Then for any $A \in \mathbb{C}^{n\times n}$ there exists a unique polynomial p over diagonal matrices of degree $n-1$ at most such that $p(P) = A$. With this representation, the problem of factoring A into the product of circulant and diagonal matrices converts into the problem of factoring p into linear factors. For a generic matrix this is possible; see Theorem 4.3. Quite intriguingly, this allows regarding matrices as polynomials which have been factored. In particular, a linear factor is, generically, the product of two diagonal matrices and a circulant matrix. Consequently, once this factoring process has been completed, we have

$$A = D_1 C_2 D_3 \cdots D_{2n-3} C_{2n-2} D_{2n-1} \tag{1.2}$$

with diagonal and circulant matrices D_{2j-1} and C_{2j}.

The paper is organized as follows. Section 2 is concerned with the set of sums of two PD matrices. Their inversion is considered. A link with the so-called $\mathcal{DCD}$ matrices is established. In Section 3, polynomials in a permutation matrix over diagonal matrices are introduced, to extend the set of the set of sums of two PD matrices. Section 4 is concerned with factoring polynomials in a permutation over diagonal matrices into first degree factors. Factorization algorithms are devised. A solution to the problem of factoring into the product of circulant and diagonal matrices is provided.

2. The sum of two PD matrices. This section is concerned with extending diagonal matrices to PD matrices, the set of scaled permutations $\mathcal{PD}$. Once done, we consider matrices consisting of the sum of two PD matrices. Here $\mathcal{P}$ denotes the set of permutations and $\mathcal{D}$ the set of diagonal matrices. In the invertible case we are dealing with the following classical matrix group.

Definition 2.1. *By monomial matrices is meant the group consisting of matrix products of permutation matrices with nonsingular diagonal matrices.*

The group property is based on the fact that if P is a permutation and D a diagonal matrix, then

$$DP = PD^P, \tag{2.1}$$

where $D^P = P^T DP$ is a diagonal matrix as well. It turns out that this "structural" commutativity allows doing practically everything the usual commutativity does. In applications, monomial matrices appear in representation theory [5, 17] and in numerical analysis of scaling and reordering linear equations [9]. See also [6, Chapter 5.3] for a link with circulant matrices. It is noteworthy that the monomial group is maximal in the general linear group of $\mathbb{C}^{n\times n}$ [8]. The following underscores that we are dealing with a natural extension of diagonal matrices.

Definition 2.2. *[1] A generalized diagonal of $A \in \mathbb{C}^{n\times n}$ is obtained by retaining exactly one entry from each row and each column of A.*

[2] The cyclic shift of size n-by-n has ones below the main diagonal and at the position $(1, n)$.

To put this into perspective in view of normality, observe that $\mathcal{PD}$ is closed under taking the Hermitian transpose. Thereby, conforming with Definition 2.2, its unitary orbit

$$\left\{U\mathcal{PD}U^* \,\middle|\, UU^* = I\right\} \tag{2.2}$$

leads to the respective notion of generalized normality. This is supported by the fact that, like for normal matrices, the eigenvalue problem for PD matrices can be regarded as being completely understood; see [6, Chapter 5.3]. To actually recover whether a given matrix $A \in \mathbb{C}^{n\times n}$ belongs to (2.2), compute the singular value decomposition $A = U\Sigma V^*$ of A and look at V^*U.[3]

PD matrices can be regarded as belonging to the more general sparse matrix hierarchy defined as follows.

Definition 2.3. *A matrix subspace $\mathcal{V}$ of $\mathbb{C}^{n\times n}$ is said to be standard if it has a basis of consisting standard basis matrices.*[4]

There is a link with graph theory. That is, standard matrix subspaces of $\mathbb{C}^{n\times n}$ are naturally associated with the adjacency matrices of digraphs with n vertices.

The following bears close resemblance to complete matching, underscoring the importance of PD matrices in linear algebra more generally through the determinant. A matrix subspace is said to be nonsingular if it contains invertible elements.

Proposition 2.4. *A matrix subspace $\mathcal{V}$ of $\mathbb{C}^{n\times n}$ is nonsingular if and only if its sparsity pattern contains a monomial matrix.*

Proof. If $A \in \mathbb{C}^{n\times n}$ is invertible, then by expanding the determinant using the Leibniz formula, one term in the sum is necessarily nonzero. The term corresponds to a monomial matrix. □

Let us now focus on the sum of two PD matrices. A monomial matrix is readily inverted by separately inverting the factors of the product. For the sum of two PD matrices, a rapid application of the inverse is also possible, albeit with different standard techniques.

Proposition 2.5. *Suppose a nonsingular $A \in \mathbb{C}^{n\times n}$ is the sum of two PD matrices. Computing a partially pivoted LU factorization of A costs $O(n)$ operations and requires $O(n)$ storage.*

Proof. Any row operation in the Gaussian elimination removes one and brings one element to the row which is being operated. Performing a permutation of rows does not change this fact. Thus, in U there are two elements in each row at most. By the symmetry, there are at most two elements in each column of L. □

Monomial matrices have a block analogue. By a block monomial matrix we mean a nonsingular matrix consisting of a permutation matrix which has in place of ones nonsingular matrices of the same size. Zeros are replaced with block zero matrices of the same size. By similar arguments, Proposition 2.5 has an analogue for the sum of two block PD matrices.[5]

The set of sums of two PD matrices, denoted by $\mathcal{PD} + \mathcal{PD}$, is no longer a group. We argue that is has many fundamental properties, though.

[3] This approach certainly works in the generic case of D having differing diagonal entries in the absolute value. In this paper we do not consider the numerical recovering of whether A belongs to (2.2) in general.

[4] A standard basis matrix of $\mathbb{C}^{n\times n}$ has exactly one entry equaling one while its other entries are zeros.

[5] Block diagonal matrices are used, e.g., in preconditioning. Thereby the sum of two block PD matrices is certainly of interest by providing a more flexible preconditioning structure.

Proposition 2.6. *$\mathcal{PD}+\mathcal{PD}$ is closed in $\mathbb{C}^{n\times n}$. Moreover, any $A\in\mathbb{C}^{n\times n}$ is similar to an element of $\mathcal{PD}+\mathcal{PD}$.*

Proof. With fixed permutations P_1 and P_2, the matrix subspace

$$\mathcal{V}=\mathcal{D}P_1+\mathcal{D}P_2. \tag{2.3}$$

is closed. Being a finite union of closed sets (when P_1 and P_2 vary among permutations), the set $\mathcal{PD}+\mathcal{PD}$ is closed as well.

For the claim concerning similarity, it suffices to observe that $\mathcal{PD}+\mathcal{PD}$ contains Jordan matrices. □

Suppose $A\in\mathbb{C}^{n\times n}$ is large and sparse. The problem of approximating A with an element of $\mathcal{PD}+\mathcal{PD}$ is connected with preprocessing. There the aim is at finding two monomial matrices so as to make $D_1P_1AD_2P_2$ more banded than A; see, e.g., [7], [4] and [3, p.441].[6] Now the permutations P_1 and P_2 in should be picked in such a way that a good approximation to A in (2.3) exists. The reason for this becomes apparent in connection with Theorem 2.7 below.

We have a good understanding of the singular elements of the matrix subspace (2.3). To see this, recall that two matrix subspaces $\mathcal{V}$ and $\mathcal{W}$ are said to be equivalent if there exist nonsingular matrices $X,Y\in\mathbb{C}^{n\times n}$ such that $\mathcal{W}=X\mathcal{V}Y^{-1}$. This is a fundamental notion. In particular, if X and Y can be chosen among permutations, then $\mathcal{V}$ and $\mathcal{W}$ are said to be permutation equivalent. In what follows, by the cyclic shift we mean the permutation

$$S=\begin{bmatrix} 0 & 0 & 0 & \cdots & 1\\ 1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \cdots & \vdots\\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{2.4}$$

of unspecified size. When $n=1$ we agree that $S=I$. The following result, which turns out to be of central relevance in extending $\mathcal{PD}+\mathcal{PD}$, should be contrasted with $(0,1)$-matrices whose line sum equals 2; see [2, Chapter 1]. Observe that, due to (2.1), $\mathcal{PD}+\mathcal{PD}$ is invariant under permutation equivalence.

Theorem 2.7. *Let $\mathcal{V}$ be the matrix subspace defined in* (2.3)*. Then*

$$\mathcal{V}=\hat{P}_1(\mathcal{D}+\mathcal{D}P)\hat{P}_2 \tag{2.5}$$

for permutations $\hat{P}_1$, $\hat{P}_2$ and $P=S_1\oplus\cdots\oplus S_k$, where S_j denotes a cyclic shift of unspecified size for $j=1,\ldots,k$.

Proof. Start by performing the permutation equivalence

$$\mathcal{V}P_2^T=\{\mathcal{D}P_1P_2^T+\mathcal{D}\}.$$

Then there are cycles associated with the matrix subspace $\mathcal{V}P_2^T$ once we represent $P_1P_2^T$ by its cycles as $P_1P_2^T=QPQ^T$ with a permutation Q. Thereby $\mathcal{V}=Q\{\mathcal{D}P+\mathcal{D}\}Q^TP_2$. □

Regarding preprocessing, observe that $\mathcal{D}+\mathcal{D}P$ in (2.5) can be regarded as essentially possessing a banded structure.

[6] The aim of preprocessing depends, to some degree, on whether one uses iterative methods or sparse direct methods; see [3, p.438].

The dimension of (2.5) is $2n$ if and only if all the cyclic shifts are of size larger than one. These matrix subspaces are sparse which is instrumental for large scale computations. In particular, it is a natural question to ask how many permutations a matrix subspace with a prescribed sparsity pattern contains. It reflects the minimum number of terms in the Leibnitz formula for determinants; see Proposition 2.4. As two extremes, in $P\mathcal{D}$ with a fixed permutation P, there is just one. And, of course, in $\mathbb{C}^{n\times n}$ there are $n!$ permutations.

COROLLARY 2.8. *There are 2^l permutations in* (2.3)*, where l is the number of cyclic shifts in* (2.5) *of size larger than one.*

Proof. The problem is invariant under a permutation equivalence, i.e., we may equally well consider $\mathcal{D}+\mathcal{D}P$. Let $\hat{P}\in\mathcal{W}$ be a permutation. When there is a cyclic shift of size one, $\hat{P}$ must have the corresponding diagonal entry. Consider the case when the cyclic shift S_j is of size larger than one. Each row and column of $\mathcal{W}$ contains exactly two nonzero elements, i.e., we must consider $\mathcal{D}+\mathcal{D}S_j$. There, by exclusion principle, $\hat{P}$ coincides either with S_j or the unit diagonal. Since $\hat{P}$ can be chosen either way, the claim follows. ▯

In general, determining the singular elements of a matrix subspace is a tremendous challenge already when the dimension exceeds two [18].[7] By using the equivalence (2.5), the singular elements of $\mathcal{V}$ can be readily determined as follows. If $D_1 = \operatorname{diag}(z_1,z_2,\ldots,z_{k_j})$ and $D_2=\operatorname{diag}(z_{k_j+1},z_{k_j+2},\ldots,z_{2k_j})$, the task consists of finding the zeros of the multivariate polynomial

$$p_j(z_1,z_2,\ldots,z_{2k_j}) = \det(D_1+D_2S_j) = \prod_{l=1}^{k_j} z_l + (-1)^{k_j-1}\prod_{l=k_j+1}^{2k_j} z_l, \tag{2.6}$$

i.e., having $\prod_{l=1}^{k_j} z_l = (-1)^{k_j}\prod_{l=k_j+1}^{2k_j} z_l$ corresponds to a singular block.

Consider a nonsingular block $D_1+D_2S_j$ under the assumption that the first (equivalently, the second) term in (2.6) is nonzero. Then its inverse can be given in a closed form with the help of the following result.

THEOREM 2.9. *Assume $S\in\mathbb{C}^{n\times n}$ is the cyclic shift and $D=\operatorname{diag}(d_1,\ldots,d_n)$. If $I+DS$ is nonsingular, then $(I+DS)^{-1}=\sum_{j=0}^{n-1}D_jS^j$ with the diagonal matrices $D_0=\frac{1}{(-1)^{n-1}\prod_{j=1}^{n}d_j+1}I$ and*

$$D_{j+1} = (-1)^{j+1}D_0\prod_{k=0}^{j} D^{S^{kT}} \quad \textit{for } j=0,\ldots,n-2. \tag{2.7}$$

Proof. It is clear that the claimed expansion exists since any matrix $A\in\mathbb{C}^{n\times n}$ can be expressed uniquely as the sum

$$A=\sum_{j=0}^{n-1}D_jS^j, \tag{2.8}$$

i.e., the diagonal matrices D_j are uniquely determined. To recover the diagonal matrices of the claim for the inverse, consider the identity

$$(I+DS)\sum_{j=0}^{n-1}D_jS^j = \sum_{j=0}^{n-1}D_jS^j + \sum_{j=0}^{n-1}DD_j^{S^T}S^{j+1} = I,$$

[7]When the dimension is two, one essentially deals with a generalized eigenvalue problem. For solving generalized eigenvalue problems there are reliable numerical methods.

where we denote SD_jS^T by $D_j^{S^T}$ as in (2.1). The problem separates permutationwise, yielding $D_0 + DD_{n-1}^{S^T} = I$ for the main diagonal and the recursion

$$D_{j+1} + DD_j^{S^T} = 0 \text{ for } j = 0, \ldots, n-2 \tag{2.9}$$

otherwise. This can be explicitly solved for $D_0 = ((-1)^{n-1}(DS)^n + I)^{-1}$. Thereby D_0 is the claimed translation of the identity matrix. Thereafter we may insert this into the recursion (2.9) to have the claim. ∎

If actually both terms on the right-hand side in (2.6) are nonzero, i.e., we are dealing with the sum of two monomial matrices, then we have a so-called $\mathcal{DCD}$ matrix, where $\mathcal{C}$ denotes the set of circulant matrices. For applications, see [14, 10] how such matrices appear in diffractive optics.

Theorem 2.10. *Assume $D_1 + D_2S$, where $S \in \mathbb{C}^{n\times n}$ is the cyclic shift and D_0 and D_1 are invertible diagonal matrices. Then there exist diagonal matrices $\hat{D}_1$ and $\hat{D}_2$ such that*

$$D_0 + D_1S = \hat{D}_1(I + \alpha S)\hat{D}_2 \tag{2.10}$$

for a nonzero $\alpha \in \mathbb{C}$.

Proof. Clearly, by using (2.1), we may conclude that the left-hand side is of more general type, including all the matrices of the type given on the right-hand side. Suppose therefore that $D_0 = \operatorname{diag}(a_1, a_2, \ldots, a_n)$ and $D_1 = \operatorname{diag}(b_1, b_2, \ldots, b_n)$ are given. Denote the variables by $\hat{D}_1 = \operatorname{diag}(x_1, x_2, \ldots, x_n)$ and $\hat{D}_2 = \operatorname{diag}(y_1, y_2, \ldots, y_n)$. Imposing the identity (2.10) yields us the equations

$$\left\{\begin{array}{ccc} x_1y_1 & = & a_1 \\ x_2y_2 & = & a_2 \\ & \vdots & \\ x_{n-1}y_{n-1} & = & a_{n-1} \\ x_ny_n & = & a_n \end{array}\right. \quad \text{and} \quad \left\{\begin{array}{ccc} \alpha x_2y_1 & = & b_1 \\ \alpha x_3y_2 & = & b_2 \\ & \vdots & \\ \alpha x_ny_{n-1} & = & b_{n-1} \\ \alpha x_1y_n & = & b_n \end{array}\right. .$$

Solving y_j in terms of x_j from the first set of equations and inserting them into the second one yields the condition $\alpha^n = \frac{\prod_{j=1}^n b_j}{\prod_{j=1}^n a_j}$ for the parameter α to satisfy. This is necessary and sufficient for the existence of a solution, obtained now by a straightforward substitution process once, e.g., the value of x_1 has been assigned. ∎

The existence of factoring (2.10) can hence be generically guaranteed in the following sense.

Corollary 2.11. *$\mathcal{D}(I + \mathbb{C}S)\mathcal{D}$ contains an open dense subset of $\mathcal{D} + \mathcal{D}S$.*

Consider the equivalence (2.5). In a generic case, using (2.10) with the blocks yields the simplest way to compute the inverse of the sum of two PD matrices.

3. Extending the sum of two PD matrices: polynomials in permutation matrices over diagonal matrices. By the fact that we have a good understanding of matrices representable as the sum of two PD matrices, we aim at extending this structure. The equivalence (2.5) provides an appropriate starting point to this end. There the canonical form consists of first degree polynomials in a permutation matrix P over diagonal matrices. More generally, define polynomials over the ring $\mathcal{D}$ with the indeterminate being an element of $\mathcal{P}$ as follows.

Definition 3.1. *Let P be a permutation and $D_k \in \mathcal{D}$ for $k = 0, 1, \ldots, j$. Then*

$$p(P) = \sum_{k=0}^{j} D_kP^k, \tag{3.1}$$

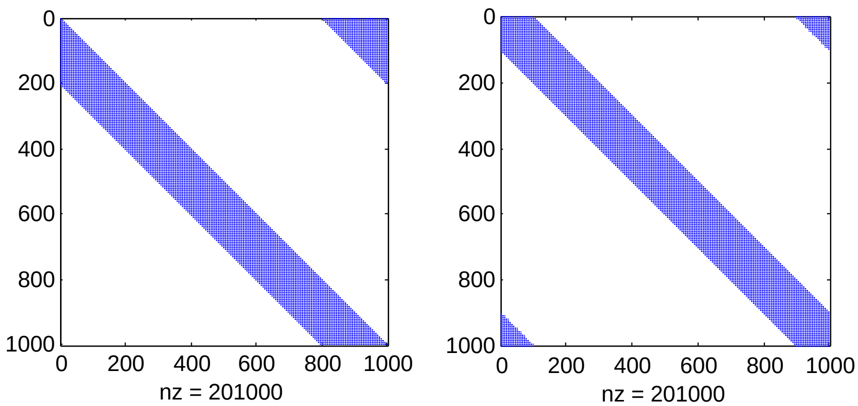

FIG. 3.1. *On the left the sparsity pattern in* (3.2) *corresponding to* $P = S$, $P_1 = P_2 = I$ *for* $n = 10^3$ *and* $j = 200$. *On the right the corresponding symmetric sparsity pattern.*

is said to be a polynomial in P *over* $\mathcal{D}$.

Due to (2.1), in terms of this representation these matrices behave in essence like standard polynomials. To avoid redundancies, we are interested in polynomials p whose degree does not exceed $\deg(P)$. Then the degree of the matrix $p(P)$ is defined to be the degree of p. For algebraic operations, the sum of polynomials $p_1(P)$ and $p_2(P)$ is obvious. Whenever $\deg p_1 + \deg p_2 < \deg(P)$, the product behaves essentially classically, i.e., the degree of the product is the sum of the degrees of the factors.

Again, bearing in mind the equivalence (2.5), there is a need to relax Definition 3.1. For this purpose, take two permutations P_1 and P_2 and consider matrix subspaces of the form

$$P_1 \left\{ p(P) \,\middle|\, \deg(p) \leq j \right\} P_2. \tag{3.2}$$

Since P_1 and P_2 can be chosen freely, by using (2.1) and (2.5) we may assume that $P = S_1 \oplus \cdots \oplus S_k$ with cyclic shifts $S_1, \ldots, S_k$. Consequently, the degrees of freedom lie in the choices of P_1 and P_2 and in the lengths of the cycles and j. Observe that (2.3) is covered by the case $j = 1$. Moreover, the sparsity structure can be made symmetric when j is even by choosing $P_1 = P^{\frac{j}{2}T}$ and $P_2 = I$. (This sparsity structure obviously contains band matrices of bandwidth $j + 1$.) This gives rise to the respective notion of "bandwidth"; see Figure 3.1.

Let us make some related graph theoretical remarks. It is natural to identify the sparsity structure of (3.2) with the $(0,1)$-matrix having the same sparsity structure.[8] Namely, there are many decomposition results to express a $(0,1)$-matrix as the sum of permutation matrices; see [2]. In this area of combinatorial matrix theory, we are not aware of any polynomial expressions of type (3.2).[9] In particular, it does not appear straightforward to see when a $(0,1)$-matrix is a realization of such a polynomial

[8]Since the study of matrix subspaces is operator space theory [15], this provides a link between analysis and discrete mathematics.

[9]It would be tempting to call such a $(0,1)$-matrix a polynomial digraph. It has, however, another meaning [2, p. 157].

structure. For example, by (2.8) we know that the matrix of all ones is. In particular, for any sparse standard matrix subspace this leads to the following notion of "graph bandwidth" in accordance with regular graphs.

DEFINITION 3.2. *Let $\mathcal{V}$ be a standard matrix subspace of $\mathbb{C}^{n\times n}$. The polynomial permutation degree of $\mathcal{V}$ is the smallest possible j allowing $\mathcal{V}$ to be included in* (3.2) *for permutations P, P_1 and P_2.*

Clearly, the polynomial degree is at most $n-1$.

The prescribed polynomial structure arises in connection with finite difference matrices with small values of j.

EXAMPLE 1. The set of tridiagonal matrices (and any of their permutation equivalences) is a matrix subspace of polynomial degree two. To see this, let P be the cyclic shift and set $j=2$, $P_1=P^T$ and $P_2=I$. Then $\mathcal{V}$ includes tridiagonal matrices. In this manner, finite difference matrices including periodic problems [9, p.159] are covered by the structure (3.2).

4. Factoring polynomials in a permutation matrix over diagonal matrices. To demonstrate that the structure (3.2) extending $\mathcal{PD}+\mathcal{PD}$ is genuinely polynomial, we want perform factoring. In forming products, we are concerned with the following algebraic structure.

DEFINITION 4.1. *Suppose $\mathcal{V}_1$ and $\mathcal{V}_2$ are matrix subspaces of $\mathbb{C}^{n\times n}$ over $\mathbb{C}$ (or $\mathbb{R}$). Then*

$$\mathcal{V}_1\mathcal{V}_2=\left\{V_1V_2 \,\middle|\, V_1\in\mathcal{V}_1 \ \textit{and}\ V_2\in\mathcal{V}_2\right\}$$

is said to be the set of products of $\mathcal{V}_1$ and $\mathcal{V}_2$.

A matrix subspace $\mathcal{V}$ is said to be factorizable if, for some matrix subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$, there holds

$$\overline{\mathcal{V}_1\mathcal{V}_2}=\mathcal{V}, \tag{4.1}$$

i.e., the closure of $\mathcal{V}_1\mathcal{V}_2$ equals $\mathcal{V}$, assuming the dimensions satisfy $1<\dim\mathcal{V}_j<\dim\mathcal{V}$ for $j=1,2$. As illustrated by the Gaussian elimination applied to band matrices, taking the closure may be necessary. For a wealth of information on computational issues related with band matrices, see [9, Chapter 4.3]. For the geometry of the set of products more generally, see [11].

Factoring in the case $j=2$ in (3.2) is handled as follows.

EXAMPLE 2. This is Example 1 continued. Let $\mathcal{V}_1=\mathcal{D}+\mathcal{D}P$ and $\mathcal{V}_2=\mathcal{D}+\mathcal{D}P^T$. Then (4.1) holds. Namely, to factor an element in a generic case, the problem reduces into solving a system of equations of the form

$$\left\{\begin{array}{lcl} x_1+\frac{a_1}{x_n} & = & b_1\\ x_2+\frac{a_2}{x_1} & = & b_2\\ x_3+\frac{a_3}{x_2} & = & b_3\\ \vdots & & \vdots\\ x_n+\frac{a_n}{x_{n-1}} & = & b_n \end{array}\right. \tag{4.2}$$

with $a_j\neq 0$ and $b_j\neq 0$ for $j=1,\ldots,n$ given. From the first equation x_1 can be solved in terms of x_n and substituted into the second equation. Thereafter x_2 can be solved in terms of x_n and substituted into the third equation. Repeating this, the system eventually turns into a univariate polynomial in x_n. Solving this combined with back substitution yields a solution. Computationally a more practical approach is to execute Newton's method on (4.2). Solving linear systems at each step is inexpensive by implementing the method of Proposition 2.5. Consequently, under standard assumptions on the convergence of Newton's method, finding a factorization is an $O(n)$ computation.

With these preparations, consider the problem of factoring a matrix subspace (3.2) into the product of lower degree factors of the same type. As described, it suffices to consider factoring a given polynomial p of degree $j \leq n-1$ in a cyclic shift $S \in \mathbb{C}^{n \times n}$ into linear factors. That is, assume having

$$p(S) = \sum_{k=0}^{j} F_k S^k \tag{4.3}$$

with diagonal matrices F_k given, for $k = 0, \ldots, j$. Then the task is to find diagonal matrices D_0 and D_1 and $E_0, \ldots, E_{j-1}$ such that

$$(D_0 + D_1 S) \sum_{k=0}^{j-1} E_k S^k = \sum_{k=0}^{j} F_k S^k \tag{4.4}$$

holds. This can then be repeated. To this end, there are several ways to proceed. Certainly, by using (2.1), the problem separates into $D_0 E_0 = F_0$ and $D_1 E_{j-1}^{S^T} = F_j$ and

$$D_0 E_{k+1} + D_1 E_k^{S^T} = F_{k+1} \tag{4.5}$$

for $k = 0, \ldots, j-2$.

There are, however, redundancies. These can be removed so as to attain maximal simplicity in terms of a univariate polynomial-like factorization result. In order to formulate a precise statement for performing this, let us invoke the following lemma.

Lemma 4.2. *Let $f : \mathbb{C}^n \to \mathbb{C}^k$ be a polynomial function. If there exists a point $x \in \mathbb{C}^n$ such that the derivative $Df(x)$ has full rank, then $f(\mathbb{C}^n)$ contains an open set whose complement is of zero measure. In particular, the open set is dense and $f(\mathbb{C}^n)$ contains almost all points of $\mathbb{C}^k$ (in the sense of Lebesgue-measure.)*

Proof. This follows from [13, Theorem 10.2]. □

Theorem 4.3. *There exists an open dense set $G \subset \mathbb{C}^{n \times n}$ containing almost all matrices of $\mathbb{C}^{n \times n}$ (in the sense of Lebesgue-measure) such that if $A \in G$, then*

$$A = (S - D_1)(S - D_2) \cdots (S - D_{n-1}) D_n \tag{4.6}$$

for diagonal matrices D_i, $i = 1, \ldots, n$.

Proof. For $1 \leq j \leq n$, define the following nj-dimensional subspaces of $\mathbb{C}^{n \times n}$

$$\mathcal{A}_j = \left\{ A \in \mathbb{C}^{n \times n} \,\middle|\, A = \sum_{k=0}^{j-1} E_k S^k \text{ for some diagonal } E_k \in \mathbb{C}^{n \times n} \right\}.$$

Consider the polynomial functions $f_j : \mathcal{A}_1 \times \mathcal{A}_{j-1} \to \mathcal{A}_j$ defined by

$$f_j(D, E) = (S - D)E.$$

After differentiating, we have

$$Df_j(D, E)(\Delta D, \Delta E) = (S - D)(\Delta E) + (-\Delta D)E.$$

Now choose $D = 0, E = I$ to obtain

$$Df_j(0, I)(\Delta D, \Delta E) = S(\Delta E) - \Delta D.$$

Hence $Df_j(0, I)$ is of full rank. By Lemma 4.2 it follows that the equation

$$f_j(D, E) = F$$

is solvable for D and E for almost all matrices $F \in \mathcal{A}_j$. Denote the subset of those matrices F by $\mathcal{B}_j = f_j(\mathcal{A}_1 \times \mathcal{A}_{j-1})$. Define $\widetilde{\mathcal{B}}_2 = \mathcal{B}_2$ and, furthermore, define

$$\widetilde{\mathcal{B}}_j = \mathcal{B}_j \cap f_j(\mathcal{A}_1 \times \widetilde{\mathcal{B}}_{j-1}), \qquad j = 3, \ldots, n.$$

Then $\mathcal{A}_j \setminus \widetilde{\mathcal{B}}_j$ is of measure zero (in $\mathcal{A}_j$) and it follows that when $A \in \widetilde{\mathcal{B}}_n$ we can solve for $D_1, \ldots, D_n$ in (4.6) by successively solving the equations (where $E_1 = A$)

$$f_j(D_j, E_{j+1}) = E_j, \qquad j = 1, 2, \ldots, n-1$$

and finally setting $D_n = E_n$. Hence almost all matrices $A \in \mathbb{C}^{n\times n}$ have a factorization (4.6). That the set of these matrices contains an open set with complement of zero measure follows by applying [13, Theorem 10.2]. ∎

The identity (4.6) allows regarding matrices as polynomials which have been factored. With these polynomials the indeterminate is a permutation (now S) while the role of $\mathbb{C}$ is taken by $\mathcal{D}$. Moreover, the representation is optimal in the sense that the number of factors (and diagonal matrices) cannot be reduced further in general. Of course, if $D_k = \alpha_k I$ with $\alpha_k \in \mathbb{C}$, then we are dealing with circulant matrices, a classical polynomial structure among matrices [6].

Like with polynomials, this gives rise to a notion of degree.

DEFINITION 4.4. *The polynomial permutation degree of $A \in \mathbb{C}^{n\times n}$ is the smallest possible j admitting a representation $A = P_1 \sum_{k=0}^{j} D_k P^k P_2$ for permutations P, P_1 and P_2 and diagonal matrices D_k for $k = 0, \ldots, j$.*

To compute the diagonal matrices D_i in (4.6) for a matrix $A \in \mathbb{C}^{n\times n}$, the equations (4.4) hence simplify as follows. Let $j = n-1$ and $A = \sum_{k=0}^{j} F_k S^k$, where F_k are diagonal. For an integer i, define $[i] = 1 + ((i-1) \bmod n)$. Denote $D_{n-j} = \operatorname{diag}(x_1, x_2, \ldots, x_n)$. Then eliminating the diagonal matrices E_k by imposing

$$(S - D_{n-j}) \sum_{k=0}^{j-1} E_k S^k = A \tag{4.7}$$

we obtain the following system of polynomial equations

$$\begin{array}{lllll}
a_{[1],1} + & a_{[2],1}x_{[1]} & + \quad a_{[3],1}x_{[1]}x_{[2]} & + \cdots + \quad a_{[j+1],1}x_{[1]}x_{[2]}\cdots x_{[j]} & = 0,\\
a_{[2],2} + & a_{[3],2}x_{[2]} & + \quad a_{[4],2}x_{[2]}x_{[3]} & + \cdots + \quad a_{[j+2],2}x_{[2]}x_{[3]}\cdots x_{[j+1]} & = 0,\\
 & & \vdots & & \\
a_{[n],n} + & a_{[n+1],n}x_{[n]} & + \; a_{[n+2],n}x_{[n]}x_{[n+1]} & + \cdots + a_{[j+n],n}x_{[n]}x_{[n+1]}\cdots x_{[n+j-1]} & = 0.
\end{array}$$

After this system has been solved, the diagonal matrices E_k can be computed by the substitutions

$$\begin{aligned}
E_{j-1} &= F_j^S,\\
E_k &= (F_{k+1} + D_{n-j}E_{k+1})^S, \qquad k = j-2, j-3, \ldots, 0.
\end{aligned}$$

We can then let $A = \sum_{k=0}^{j-1} E_k S^k$, decrease j by one and repeat the solving of (4.7) accordingly.

Let us now return to our original problem of factoring into the product of circulant and diagonal matrices. Certainly, Theorem 2.10 can be combined with Theorem 4.3 to have a factorization after completing the prescribed computations. For another approach, to directly factor a matrix $A = p(S)$ into the product of circulant and diagonal matrices, the following approach allows ignoring E_k's completely. Namely, assuming D_0 and D_1 to be invertible, use Theorem 2.10 to have

$$\sum_{k=0}^{j-1} \hat{E}_k S^k = (I + \alpha S)^{-1} \tilde{D}_1 p(S) \tag{4.8}$$

with $\hat{E}_k = \hat{D}_2 E_k$, $\alpha \in \mathbb{C}$ and $\tilde{D}_1 = \hat{D}_1^{-1}$. Clearly, $\hat{D}_2$ is redundant. Thereby the task reduces to choosing α and $\tilde{D}_1 = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ in such a way that the right-hand side of the identity attains the zero structure imposed by the left-hand side. Any solution is homogeneous in $\tilde{D}_1$. Therefore we can further set $d_1 = 1$ to reduce the problem to n free complex parameters. Once the equations are solved, $\hat{E}_k$'s are determined by α and $\tilde{D}_1$ without any further effort.

To factor by using (4.8), let $j = n - 1$, i.e., consider the first factorization step. Then zeros on the left-hand side of (4.8) appear at the positions where $S^{n-1} = S^T$ has ones. To have the functions on right-hand size at these positions, the inverse of $I + \alpha S$ is the circulant matrix with the first row

$$\frac{1}{1 + (-1)^{n-1}\alpha^n}(1, (-1)^{n-1}\alpha^{n-1}, (-1)^{n-2}\alpha^{n-2}, \ldots, \alpha^2, -\alpha) \tag{4.9}$$

by Theorem 2.9. In the arising polynomial equations the factor $\frac{1}{1+(-1)^{n-1}\alpha^n}$ can be ignored. (In the equations of interest, the denominator multiplies zeros.) Thereby we have n polynomial equations in which the highest power of α is $n - 1$ while d_j's appear linearly. These equations are readily written down.

Once the factorization is completed we have (1.2). The number of free parameters is $n^2 + n - 1$ by the fact that the circulant matrices C_k appearing in the factorization are of the form $I + \alpha_k S$ for $\alpha_k \in \mathbb{C}$. Hence this leaves us only $n - 1$ "excess" free parameters.

Example 3. The matrix $p(S)$ in (4.3) is doubly stochastic if $F_k = f_k I$ with $f_k \geq 0$ such that $\sum_{k=0}^{j} f_k = 1$. Regarding the degrees of freedom, it might be of interest to factor $p(S)$ into the product of doubly stochastic matrices of lower order. For the factors $I + \alpha_k S$ this can be readily done.

Let us end the paper with a speculative deliberation on the optimal number of factors. Regarding the factorization problem of a generic matrix into the minimal number of circulant and diagonal factors, we make the following conjecture.

Conjecture 1. There exists an open dense set $G \subset \mathbb{C}^{n\times n}$ containing almost all matrices of $\mathbb{C}^{n\times n}$ (in the sense of Lebesgue-measure) such that if $A \in G$, then

$$A = B_1 B_2 \cdots B_{n+1},$$

where $B_i \in \mathbb{C}^{n\times n}$ is circulant for odd i and diagonal for even i.

This is supported by calculations. That is, we have verified the conjecture for the dimensions n satisfying $2 \leq n \leq 20$ by computer calculations utilizing Lemma

4.2 (with randomly chosen integer coordinates for the point x resulting in an integer matrix for the derivative). Observe that, by a simple count of free parameters, no lower number of factors can suffice.

REFERENCES

[1] R. Bhatia, *Trimming, truncating and averaging of matrices,* Amer. Math. Monthly, 107 (2000), pp. 602–608.

[2] R. Brualdi and H. Ryser, Combinatorial Matrix Theory, *Cambridge University Press, Cambridge,* 1991.

[3] M. Benzi, *Preconditioning techniques for large linear systems: a survey,* J. Comput. Phys., 182 (2002), no. 2, pp. 418–477.

[4] M. Benzi, J. C. Haws and M. Tůma, *Preconditioning highly indefinite and nonsymmetric matrices*, SIAM J. Sci. Comput., 22 (4) (2000), pp. 1333–1353.

[5] C.W. Curtis and I. Reiner, Representation Theory of Finite Groups and Associative Algebras, *AMS Chelsea Publishing*, 1962.

[6] P. Davis, Circulant Matrices, *Wiley, New York,* 1979.

[7] I.S. Duff and J. Koster, *The design and use of algorithms for permuting large entries to the diagonal of sparse matrices*, SIAM J. Matrix Anal. and Appl., 20 (1999), pp. 889–901.

[8] S. Friedland, *Maximality of the monomial group,* Lin. Multilin. Algebra, 18 (1985), pp. 1–7.

[9] G.H. Golub and C.F. van Loan, Matrix Computations, *The Johns Hopkins University Press, the 3rd ed.*, 1996.

[10] M. Huhtanen, *Approximating ideal diffractive optical systems,* J. Math. Anal. Appl., 345 (2008), pp. 53–62.

[11] M. Huhtanen, *The product of matrix subspaces*, arXiv:1111.7143.

[12] T. Laffey, *Conjugacy and factorization results on matrix groups,* Functional analysis and operator theory (Warsaw, 1992), Banach Center Publ., 30, Polish Acad. Sci., Warsaw, (1994) pp. 203–221.

[13] J. S. Milne, Algebraic Geometry, *www.jmilne.org/math/,* 2012.

[14] J. Müller-Quade, H. Aagedal, T. Beth and M. Schmid, *Algorithmic design of diffractive optical systems for information processing,* Physica D, 120 (1998) pp. 196-205.

[15] V. Paulsen, Completely Bounded Maps and Operator Algebras, *Cambridge University Press, Cambridge,* 2002.

[16] M. Schmid, R. Steinwandt, J. Müller-Quade, M. Rötteler and T. Beth, *Decomposing a matrix into circulant and diagonal factors,* Linear Algebra Appl., 306 (2000), pp. 131–143.

[17] J.J. Rotman, An Introduction to the Theory of Groups, *Springer-Verlag, New York*, 1994.

[18] O. Taussky, *The characteristic polynomial and the characteristic curve of a pencil of matrices with complex entries,* Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber., II 195 (1986), no. 1-3, pp. 175–178.