Decay in Functions of Multiband Matrices

Mastronardi, N.; Ng, M.; Tyrtyshnikov, E. E.

Published in:
SIAM Journal on Matrix Analysis and Applications

DOI:
10.1137/090758374

Published: 07/10/2010

SIAM J. MATRIX ANAL. APPL.
Vol. 31, No. 5, pp. 2721–2737

DECAY IN FUNCTIONS OF MULTIBAND MATRICES*

N. MASTRONARDI†, M. NG‡, AND E. E. TYRTYSHNIKOV§

Abstract. The Benzi–Golub result on decay properties for matrix functions of a banded Hermitian matrix [*BIT*, 39 (1999), pp. 417–438] is extended to the case of multiband matrices. It is shown how the simple diagonal dominance technique applies to the general non-Hermitian case. We also present $O(1)$ algorithms computing matrix functions of multiband and multi-Toeplitz (multilevel Toeplitz) matrices in time that depends on the bandwidth and prescribed approximation accuracy but does not depend on the size of matrices.

Key words. matrix functions, banded matrices, multilevel matrices, multiband matrices, polynomial approximation, exponential decay, Toeplitz matrices, numerical range

AMS subject classifications. 65F30, 65F60, 65D15

DOI. 10.1137/090758374

1. Introduction.

Decay properties of the entries of inverses to banded diagonally dominant matrices have been known for a long time: exponential decay away from the main diagonal is granted the actual rate depending on the bandwidth [10]. The exponential decay property in some algebras and applications of operators with pseudosparse matrices was studied in [6]. Then a similar observation to [10] was established for infinite Hermitian positive definite matrices [11], with some care to be taken about the constant factor when applied to finite matrices. Furthermore, a far reaching generalization was made in [3]: the same decay properties turn out to hold for a wide class of functions of a given banded Hermitian positive definite matrix, rather than only inverses. Extensions to non-Hermitian matrices and $O(n)$ algorithms for approximating functions of sparse matrices are presented in [4]. The importance of the latter results is amplified by a conspicuous growth of interest in computations with matrix functions [1, 5, 12, 16]. More references on important related works are given in [4].

However, multidimensional problems naturally give rise to multilevel sparse matrices with a bandwidth depending on the matrix size, for instance, in Markov chains [7, 21] and in image segmentation [17]. As a consequence, the decay rates of the previous results depend on the matrix size. Nevertheless, the same matrices usually can be viewed as multilevel multiband matrices with the multiband sizes independent of the order of matrix. In this paper we propose that the decay results for such matrices

*Received by the editors May 7, 2009; accepted for publication (in revised form) July 22, 2010; published electronically October 7, 2010.

http://www.siam.org/journals/simax/31-5/75837.html

†Istituto per le Applicazioni del Calcolo "M. Picone," CNR, Sezione di Bari, 70126 Bari, Italy (n.mastronardi@ba.iac.cnr.it). This author's work was supported by a Russian-Italian collaboration agreement between RAS and CNR.

‡Centre for Mathematical Imaging and Vision and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (mng@math.hkbu.edu.hk). This author's work was supported by RGC Research Grants Council grants and HKBU FRGs.

§Institute of Numerical Mathematics, Russian Academy of Sciences, 11999, Moscow, Russia and Institute for Computational Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (tee@inm.ras.ru). This author's work was supported by the Russian Foundation for Basic Research (RFBR 08-01-00115, RFBR/DFG 09-01-91332), Priority Research Grant of the Mathematical Sciences Department of the Russian Academy of Sciences, and Russian-Italian collaboration agreement between the Russian Academy of Sciences and CNR.

can be more suitably presented in the multi-index notation, for the decay takes place at every level of a given multilevel matrix.

We use a definition of multilevel matrix suggested in [22]: a matrix A of order N is said to have p levels of sizes $n_1, \dots, n_p$ if $N = n_1 \dots n_p$, A is viewed as a block $n_1 \times n_1$ matrix with every block being considered as a block $n_2 \times n_2$ matrix, and, hierarchically, each block $n_i \times n_i$ matrix with every block being considered as a block $n_{i+1} \times n_{i+1}$ matrix for $i = 2, 3, \dots, p-1$. For instance, A has two levels of sizes n_1 and n_2, and A is just a block $n_1 \times n_1$ matrix with $n_2 \times n_2$ matrix block. The row and column indices of A are in the one-to-one correspondence with p-tuple indices:

$$i \leftrightarrow \mathbf{i} = (i_1, \dots, i_p), \quad j \leftrightarrow \mathbf{j} = (j_1, \dots, j_p),$$

where, by definition,

$$\begin{array}{rcl} i &=& (i_1 - 1)n_2 \dots n_p + (i_2 - 1)n_3 \dots n_p + \dots + (i_{p-1} - 1)n_p + i_p, \\ j &=& (j_1 - 1)n_2 \dots n_p + (j_2 - 1)n_3 \dots n_p + \dots + (j_{p-1} - 1)n_p + j_p. \end{array}$$

Thus, for the entries of A, we shall write

$$a_{ij} = a_{\mathbf{ij}} = a(i_1, \dots, i_p; j_1, \dots, j_p).$$

In this case we say that A is a matrix of multiorder $\mathbf{n} = (n_1, \dots, n_p)$.

A p-level matrix A is called multiband of multiwidth $\mathbf{m} = (m_1, \dots, m_p)$ if it is granted that

$$a(i_1, \dots, i_p; j_1, \dots, j_p) = 0$$

whenever

$$|i_l - j_l| > m_l/2$$

at least for one value of $l \in \{1, \dots, p\}$. Each number m_l is assumed to be odd and is referred to as the bandwidth of level l. It is equivalent to say that A is a block banded matrix with the block bandwidth m_1, each nonzero block is again a block banded matrix with the block bandwidth m_2, and so on.

Multiband matrices are sparse with a special sparsity pattern. This pertains to a more general type of linear structure expounded in [18].

We present certain conditions under which the entries of a p-level matrix $B = f(A)$ are estimated in the following way:

$$|b(i_1, \dots, i_p; j_1, \dots, j_p)| \le cq^{\max_l |i_l - j_l|}, \quad 0 < q < 1, \tag{1.1}$$

where c is determined by certain properties of A and f. Note that exponential decrease is ascertained in the upper estimate for the entries and that the estimate is not necessarily sharp. Usually we derive this estimate for a family of matrices of different multiorders $\mathbf{n}$. In such cases, if c does not depend on $\mathbf{n}$, then the estimate becomes useful for sufficiently large sizes of levels. For an individual matrix, it may help when the constant c is not far away from the maximal modulus of the entries of A and q. In the mentioned cases, the estimate in (1.1) is considered for the proof and construction of low-parametric approximations to typical matrix functions of A.

A principal contribution of this paper is that we provide estimates for the multiband case. However, we suggest as well some other novelties.

As in all previous works, we assume that f is analytic in a domain containing the spectrum of A. However, on a qualitative level, for non-Hermitian matrices we present the decay results by using entirely elementary tools such as the Gershgorin localization theorem and easy consequences of the diagonal dominance (Theorem 2.5). Of course, we pay for simplicity by taking up a possibly larger analyticity domain for f and eventually by increasing the value of c. All the same, even if a domain is the smallest one containing the spectrum, the estimate of [4] assumes that a non-Hermitian matrix A is diagonalizable. We propose some estimates where this assumption is not used.

In practice and in model examples, multiband matrices frequently possess one more important property: they are also multi-Toeplitz (i.e., multilevel matrices with Toeplitz structure on every level) or can be made multi-Toeplitz by changing just few entries in the corners of blocks of each level. For approximating matrix functions for this class of matrices, we propose $O(1)$ algorithms, with the number of operations depending on the bandwidth and prescribed approximation accuracy but *not* on the size of the matrix.

In order to illustrate and confirm the theoretical claims, we present some numerical experiments.

The outline of this paper is as follows. In section 2, we present some decay results. In section 3, we study the results for multi-Toeplitz matrices. In section 4, we present numerical examples. Finally, some concluding remarks are given in section 5.

2. Decay results.

Consider a sequence $S_0 \subset S_1 \subset \dots$ of nonempty subsets of

$$S = \{(i, j) : i, j = 1, 2, \dots\}.$$

For any matrix A, denote by $\mathrm{NZ}(A)$ the set of all index pairs $(i, j) \in S$ such that $a_{ij} \neq 0$. Define $||A||_C$ as the maximal in modulus entry of A.

LEMMA 2.1. *Given a matrix A, assume that* $\mathrm{NZ}(A^k) \subset S_k$ *for* $k = 0, 1, \dots,$ *and consider a matrix B and a function $E(d)$ with the following property: for any* $d = 0, 1, \dots,$ *there exists a polynomial $p_d(x)$ such that* $\deg(p_d) \le d$ *and*

$$||B - p_d(A)||_C \le E(d).$$

Then

$$|b_{ij}| \le E(d)$$

whenever

$$(i, j) \notin S_d.$$

Proof. All nonzero entries of $p_d(A)$ occupy positions $(i, j) \in S_d$. Hence, from $(i, j) \notin S_d$, we conclude that $|b_{ij}| = |(B - p_d(A))_{ij}| \le E(d)$. □

LEMMA 2.2. *Let A be a p-level multiband matrix with level bandwidths $m_1, \dots, m_p$. Then*

$$\mathrm{NZ}(A^k) \subset S_k,$$

where for all levels l,

$$S_k = \{(i, j) = (\mathbf{i}, \mathbf{j}) : |i_l - j_l| < (km_l - k + 1)/2\}.$$

Proof. It follows from the observation that the product AB of two matrices with bandwidths $m(A)$ and $m(B)$ is a band matrix with the bandwidth $m(AB) = m(A) + m(B) - 1$. □

LEMMA 2.3. *Let A be a p-level multiband matrix with level bandwidths $m_1, \dots, m_p$. Assume that a matrix B and polynomials $p_d(x)$ are such that $\deg(p_d) \le d$ and*

$$||B - p_d(A)||_C \le E(d)$$

with the monotonicity property

$$E(d+1) \le E(d), \quad d = 0, 1, \dots.$$

Then B can be considered as a p-level matrix with the same level sizes as those of A, and the entries of B outside the multiband of A are bounded as follows:

$$|b_{ij}| = |b_{\mathbf{ij}}| \le E\left(t(i,j)\right),$$

where

$$t(i,j) = \max_{1 \le l \le p} \left\lceil \frac{2|i_l - j_l|}{m_l} \right\rceil.$$

Proof. Consider a position (i, j), choose level l, and find a positive integer d such that

$$(dm_l - d + 1)/2 < |i_l - j_l| \le dm_l/2.$$

Then $(i, j) \notin S_d$, and hence,

$$|b_{ij}| \le E(d) \le E\left(\left\lceil \frac{2|i_l - j_l|}{m_l} \right\rceil\right).$$

Using the monotonicity of $E(d)$, we diminish the right-hand side by maximizing the argument. ∎

These lemmas help us to derive estimates for the entries of those matrices B that can be approximated by a polynomial of a multiband matrix A. The decay properties are determined by the behavior of approximation accuracy with regard to the degree of polynomial.

A simple polynomial approximation result for $B = (I + F)^{-1}$ is easily available in the case $||F||_2 < 1$. The latter inequality implies the convergence of the Neumann series

$$(I + F)^{-1} = \sum_{s=0}^{\infty} (-1)^s F^s.$$

Upon truncation down to $d + 1$ terms, we obtain a polynomial

$$p_d(F) = \sum_{s=0}^{d} (-1)^s F^s.$$

Then we can get the estimate

$$||B - p_d(F)||_C \le c_\rho \rho^d, \quad c_\rho = \frac{\rho}{1 - \rho},$$

where $\rho = \|F\|_2$. This leads us immediately to the following theorem.

THEOREM 2.4. *Let A be a p-level multiband matrix with level bandwidths $m_1, \ldots, m_p$, and assume that A is diagonally dominant in rows:*

$$\rho |a_{ii}| \geq \sum_{j \neq i} |a_{ij}|, \quad 1 \leq i \leq n, \quad 0 < \rho < 1.$$

Then the entries of A^{-1} outside the multiband of A satisfy the bounds

$$|(A^{-1})_{ij}| \leq c \rho^{t(i,j)}, \quad t(i,j) = \max_{1 \leq l \leq p} \left[\frac{2|i_l - j_l|}{m_l} \right],$$

with

$$c = \left(\max_i |a_{ii}^{-1}| \right) \frac{\rho}{1-\rho}.$$

Proof. Let $D = \mathrm{diag}(A)$ denote the main diagonal of A. Then $D^{-1}A = I + F$ with $||F||_2 \leq ||F||_\infty \leq \rho$. It remains to note that $A^{-1} = (I+F)^{-1} D^{-1}$ and to apply the above polynomial approximations to $(I+F)^{-1}$. ◻

Similar results can be obtained for matrices of the form $B = f(A)$, where the function f is well approximated by polynomials around the spectrum of A. A rather straightforward generalization of Theorem 2.4 is as follows.

THEOREM 2.5. *Assume that $f(z)$ is analytic in a bounded simply connected domain with a rectifiable (e.g., piecewise smooth) boundary Γ and that a matrix A and a number $0 < \rho < 1$ are such that*

$$\rho |a_{ii} - z| \geq \sum_{j \neq i} |a_{ij}|, \quad 1 \leq i \leq n, \quad z \in \Gamma.$$

Let A be a p-level multiband matrix with level bandwidths $m_1, \ldots, m_p$. Then the entries of $f(A)$ outside the multiband of A are bounded as follows:

$$|\{f(A)\}_{ij}| \leq c \rho^{t(i,j)}, \quad t(i,j) = \max_{1 \leq l \leq p} \left[\frac{2|i_l - j_l|}{m_l} \right],$$

with

$$c = \frac{2\pi\rho}{1-\rho} \max_{z \in \Gamma} \min_{1 \leq i \leq n} |a_{ii} - z|^{-1} \int_{z \in \Gamma} |f(z)| d|z|.$$

Proof. By the Gershgorin theorem, all the eigenvalues of A are strictly inside the domain encircled by Γ. Hence,

$$f(A) = \frac{1}{2\pi \mathbf{i}} \int_\Gamma (A - zI)^{-1} f(\zeta) d\zeta.$$

We observe that the matrix $A - zI$ keeps the multiband structure of A for all z. Then the claimed estimate follows by direct application of Theorem 2.4. ◻

More sophisticated approximation techniques may result in tighter estimates. Assume that f is analytic in a compact simply connected domain $\mathcal{D}$ with a rectifiable (e.g., piecewise smooth) boundary Γ, and let the spectrum of A lie in a compact domain $\mathcal{D}'$ strictly inside $\mathcal{D}$. Then, for any polynomial $p_d(z)$ of degree d or less and for any $z \in \mathcal{D}'$, we obtain

$$f(A) - p_d(A) = \frac{1}{2\pi \mathbf{i}} \int_\Gamma (A - \zeta I)^{-1} (f(\zeta) - p_d(\zeta)) d\zeta.$$

Consequently,

$$||f(A) - p_d(A)||_2 \le C_\Gamma \, R_\Gamma(A) \, ||f - p_d||_\Gamma,$$

where

$$C_\Gamma = \frac{1}{2\pi} \int\limits_\Gamma |d\zeta|, \quad R_\Gamma(A) = \max_{\zeta \in \Gamma} ||(A - \zeta I)^{-1}||_2,$$

$$||f - p_d||_\Gamma = \max_{\zeta \in \Gamma} |f(\zeta) - p_d(\zeta)|.$$

By Mergelyan's generalization of the Runge theorem, if the complement of $\mathcal{D}$ is connected, then

$$\lim_{d \to \infty} \min_{\deg p_d \le d} ||f - p_d||_\Gamma = 0. \tag{2.1}$$

Thus, the decay estimates can follow from the convergence estimates in (2.1).

It is worth noting that, in some cases, the norm of the resolvent can be easily estimated from above by a value that does not depend on A. Indeed, if D' contains the *numerical range* of A, defined as a set

$$W(A) \equiv \{(Ax, x) \in \mathbb{C} \, : \, ||x||_2 = 1\},$$

then

$$R_\Gamma(A) \le \frac{1}{\mathrm{dist}(\Gamma, \mathcal{D}')}.$$

Moreover, as is proved in [9], if $\mathcal{D}'$ is a compact convex set, then there exists a constant $C(\mathcal{D}')$ such that

$$||p_d(A)||_2 \le C(\mathcal{D}') \, ||p_d||_{\mathcal{D}'}.$$

Therefore,

$$||f(A) - p_d(A)||_2 \le C(\mathcal{D}') \, ||f - p_d||_{\mathcal{D}'}.$$

Using the Mergelyan theorem, we conclude that $f(A)$ can be approximated by polynomials of A. Note that $C(\mathcal{D}')$ does not depend on A; it is required only that $W(A) \subset \mathcal{D}'$. Even more than that, Crouzeix proved that $C(\mathcal{D}') \le 11.08$ [8], and his conjecture is that $C(\mathcal{D}') = 2$ (so the constant does not depend on $\mathcal{D}'$; note that in [9], the estimate was considered for rational functions, and dependence on $\mathcal{D}'$ was conceded). For a discussion and application of these results to the convergence analysis of GMRES, see [2] and signal processing [14].

Further, polynomial approximations $p_d(z)$ to $f(z)$ with the property

$$||f(z) - p_d(z)||_{\mathcal{D}'} \le C_f \, \rho^d, \quad 0 < \rho < 1, \tag{2.2}$$

can be constructed from the expansion of $f(z)$ in the so-called Faber polynomials (cf. [4, 5]). The construction capitalizes on a function $z = \psi(w)$ of the form

$$\psi(w) = \gamma \left(w + d_0 + \frac{d_1}{w} + \dots \right)$$

that maps the exterior of the unit disc conformally onto the exterior of $\mathcal{D}'$. The coefficient γ is defined uniquely and is known as the *logarithmic capacity* of $\mathcal{D}'$. The inverse mapping $w = \phi(z)$ to $z = \psi(w)$ is obviously of the same form,

$$\phi(z) = \frac{1}{\gamma}\left(z + c_0 + \frac{c_1}{z} + \dots\right),$$

and the Faber polynomials are defined as the polynomial parts of the Laurent expansions of the powers $(\phi(z))^k$ for $k = 1, 2, \dots$. If the exterior of $\mathcal{D}$ is obtained by the same mapping $\psi(z)$ from the exterior of a disc of radius $r > 1$, then we can take $\rho = 1/r$, and this value of ρ cannot be decreased on the whole set of functions f analytical inside $\mathcal{D}$ and continuous on $\mathcal{D}$.

All in all, we arrive at the following theorem.

THEOREM 2.6. *Let A be a p-level multiband matrix with level bandwidths $m_1, \dots, m_p$, and let the numerical range of A belong to a compact convex set $\mathcal{D}'$. Assume that $f(z)$ admits polynomial approximations on $\mathcal{D}'$ with the estimate* (2.2). *Then the entries of $f(A)$ outside the multiband of A are bounded as follows:*

$$|\{f(A)\}_{ij}| \le c\rho^{t(i,j)}, \quad t(i,j) = \max_{1\le l\le p}\left[\frac{2|i_l - j_l|}{m_l}\right], \tag{2.3}$$

with $c = C_f C(\mathcal{D}')$.

A method of evaluation of the constants C_f and ρ is simplified if $\mathcal{D}'$ and $\mathcal{D}$ have cofocal elliptic boundaries. In such cases, $p_d(z)$ can be obtained as an interpolation polynomial at the Chebyshev nodes.

Let $1 < r$ and E_r denote a compact domain bounded by an ellipse Γ with foci at the points -1 and 1 and the main axis $r + r^{-1}$. Note that Γ_r is known as the *Bernstein ellipse*. It is easy to check that the half-axes of Γ_r are

$$a = \frac{1}{2}(r + r^{-1}), \quad b = \frac{1}{2}(r - r^{-1}),$$

and that E_r dilates as r increases:

$$0 < r_1 < r_2 < 1 \quad \Rightarrow \quad E_{r_1} \subset E_{r_2}.$$

For specific values of r, the sets E_r might serve as the analyticity domain for f. Still it would be a rather big restriction, e.g., the logarithm function is not analytic on any of E_r. Moreover, in order just to embrace the spectra of some matrices, we would be caused to take up unnecessarily large domains. To be better off with practical cases, it is common to consider domains $\mathcal{D}_r$ obtained from E_r by scaling, shift, and rotation (same for all r). Thus, let us assume that the numerical range of A belongs to a domain

$$\mathcal{D}_{r'} \equiv \psi(E_{r'}),$$

where

$$\psi(w) = \xi + \eta w, \quad \xi, \eta \in \mathbb{C}, \quad \eta \ne 0.$$

The inverse mapping $w = \phi(z)$ to $z = \psi(w)$ is of the same affine form:

$$\phi(z) = \alpha + \beta z, \quad \alpha, \beta \in \mathbb{C}, \quad \beta \ne 0.$$

Let f be such that $F(w) \equiv f(\psi(w))$ is analytic inside E_r and continuous on E_r with $1 < r' < r$. Denote by $P_{d-1}(w)$ a polynomial interpolating $F(w)$ at the roots $w_1, \dots, w_d$ of the Chebyshev polynomial

$$T_d(w) = \cos(d \arccos w) = 2^{d-1}\Omega(w), \quad \Omega(w) = \prod_{j=1}^{d} (w - w_j).$$

Choose and fix any $w \in E_{r'} \backslash \{w_1, \dots, w_d\}$. Then the function

$$\frac{F(\zeta)}{\Omega(\zeta)(\zeta - w)}$$

has only simple poles, and hence (cf. [13]),

$$\frac{1}{2\pi \mathbf{i}} \int\limits_{\Gamma_r} \frac{F(\zeta)}{\Omega(\zeta)(\zeta - w)} d\zeta = \sum_{j=1}^{d} \frac{F(w_j)}{\Omega'(w_j)(w_j - w)} + \frac{F(w)}{\Omega(w)} = \frac{1}{\Omega(w)} \left(F(w) - P_{d-1}(w)\right).$$

It directly follows that

$$F(w) - P_{d-1}(w) = \frac{T_d(w)}{2\pi \mathbf{i}} \int\limits_{\Gamma_r} \frac{F(\zeta)}{T_d(\zeta)(\zeta - w)} d\zeta.$$

Using the well-known properties of Chebyshev polynomials on the Bernstein ellipses

$$\frac{1}{2}\left(r^d - r^{-d}\right) \le |T_d(w)| \le \frac{1}{2}\left(r^d + r^{-d}\right),$$

for $w \in E_{r'}$, we have the estimate

$$|F(w) - P_{d-1}(w)| \le C_f\, \rho^d, \quad \rho = \frac{r'}{r}, \tag{2.4}$$

$$C_f = \frac{r}{r-1} \, \frac{\int_{\Gamma_h} |F(\zeta)||d\zeta|}{2\pi\, \mathrm{dist}(\Gamma_r, \Gamma_{r'})}. \tag{2.5}$$

It remains to set $p_d(z) = P_d(\alpha + \beta z)$, and note that $|f(z) - p_d(z)| = |F(w) - P_d(w)|$ so long as $w = \phi(z) = \alpha + \beta z$.

This method suggests as well a fast and stable algorithm for the computation of $P_d(w)$. We can use an expansion in the Chebyshev polynomials

$$F(x) \approx P_d(x) = \sum_{l=0}^{d} s_l T_l(x), \quad -1 \le x \le 1,$$

take up the Chebyshev nodes

$$w_{k+1} = \cos\left(\frac{\pi(2k+1)}{2(d+1)}\right),$$

and consider the interpolation conditions

$$P_d(w_{k+1}) = \sum_{l=0}^{d} s_l T_l(w_{k+1}) = F(w_{k+1}), \quad 0 \le k \le d.$$

It is easy to see that s_l are determined from a system of linear algebraic equations with the coefficient matrix of the cosine transform

$$F = \left[\cos\left(\frac{\pi(2k+1)l}{2(d+1)}\right)\right], \quad 0 \le k, l \le d,$$

which is easily invertible by fast Fourier transform. However, fastness might be not so crucial as we do not intend to deal with high-order polynomials. Another important observation is that F becomes an orthogonal matrix after scaling the columns.

3. Multi-Toeplitz results. A p-level matrix A with the entries

$$a_{\mathbf{ij}} = a(i_1, \ldots, i_p; j_1, \ldots, j_p)$$

is called *multi-Toeplitz* if

$$a_{\mathbf{ij}} = a_{\mathbf{i}'\mathbf{j}'} \quad \text{whenever} \quad \mathbf{i} - \mathbf{j} = \mathbf{i}' - \mathbf{j}'.$$

For such matrices it is common to write

$$a_{\mathbf{ij}} = a_{\mathbf{i}-\mathbf{j}}.$$

Assume also that A is a multiband matrix, and inquire into the gains of the combination of multiband and multi-Toeplitz structures.

To begin, consider a one-level tridiagonal matrix A. We know already that A^k has $2k+1$ diagonals. However, if A is Toeplitz, then A^k retains the Toeplitz property in the bulk of elements, e.g., as follows:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 2 & 2 & 1 & 0 & 0 & 0 \\ 2 & 3 & 2 & 1 & 0 & 0 \\ 1 & 2 & 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 \\ 0 & 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & 1 & 2 & 2 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 4 & 5 & 3 & 1 & 0 & 0 \\ 5 & 7 & 6 & 3 & 1 & 0 \\ 3 & 6 & 7 & 6 & 3 & 1 \\ 1 & 3 & 6 & 7 & 6 & 3 \\ 0 & 1 & 3 & 6 & 7 & 5 \\ 0 & 0 & 1 & 3 & 5 & 4 \end{bmatrix}.$$

We can observe that A^2 is a Toeplitz matrix except the corner 3×3 matrix, A^3 is again a Toeplitz matrix except for the corner 4×4 matrix, and so on. Also, the framed blocks completely determine the corresponding matrices. To be more precise, in the framed blocks, it is enough to know the elements except the entries in some antidiagonals, e.g., the $(1,1)$th, $(1,2)$th, and $(2,1)$th entries of A^3 are different from the diagonals of the Toeplitz part.

Another interesting observation is that the elements in the upper part from the antidiagonal in the framed $(k+1) \times (k+1)$ block of A^k coincide with the corresponding elements of B^k, where B is the same-size leading block in A. This leads to a very simple recipe for acquisition of the defining elements of A^k in $O(1)$ operations: we need just to compute B^k, where the size of B does not depend on the size of A, and the latter can be arbitrarily large.

It is also useful to remark that A^k is symmetric with respect to the antidiagonal. Such matrices are called *persymmetric* and satisfy the equality

$$A^\top = JAJ, \quad J = \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}.$$

It is well known and easy to check that any Toeplitz matrix is persymmetric. Consequently, any polynomial of a Toeplitz matrix remains persymmetric.

THEOREM 3.1. *Let $A = A_1 \dots A_k$ be a product of one-level banded Toeplitz matrices $A_1, \dots, A_k$ of order n, each with $2m+1$ diagonals. Then $A = [a_{ij}]$ coincides with a Toeplitz matrix in the index domain*

$$2 + m(k-1) \le i + j \le 2n - m(k-1).$$

Proof. By induction, assume that the claim is established for A, and consider a matrix $C = BA$, where B is an arbitrary Toeplitz banded matrix with $2m+1$ diagonals. Obviously,

$$B = b_0 I + \sum_{l=1}^{m} \left(b_l Z^l + b_{-l} (Z')^l \right),$$

where Z is the downshift matrix,

$$Z = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}.$$

The key observation is that $Z^l B$ is Toeplitz in the index domain

$$2 + mk \le i + j \le 2n. \tag{3.1}$$

For illustration, let $n = 10$, $m = 3$, and $k = 2$. Then, symbolically,

$$A = \begin{bmatrix} * & * & * & * & t & t & t & 0 & 0 & 0 \\ * & * & * & t & t & t & t & t & 0 & 0 \\ * & * & t & t & t & t & t & t & t & 0 \\ * & t & t & t & t & t & t & t & t & t \\ t & t & t & t & t & t & t & t & t & t \\ t & t & t & t & t & t & t & t & t & t \\ t & t & t & t & t & t & t & t & t & * \\ 0 & t & t & t & t & t & t & t & t & * \\ 0 & 0 & t & t & t & t & t & t & * & * \\ 0 & 0 & 0 & t & t & t & * & * & * & * \end{bmatrix}, \quad ZA = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & * & * & * & t & t & t & 0 & 0 & 0 \\ * & * & * & t & t & t & t & t & 0 & 0 \\ * & * & t & t & t & t & t & t & t & 0 \\ * & t & t & t & t & t & t & t & t & t \\ t & t & t & t & t & t & t & t & t & t \\ t & t & t & t & t & t & t & t & t & t \\ t & t & t & t & t & t & t & t & t & * \\ 0 & t & t & t & t & t & t & t & t & * \\ 0 & 0 & t & t & t & t & t & t & * & * \end{bmatrix},$$

where t stands for the Toeplitz-part entries and $*$ for the non-Toeplitz entries. Multiplication from the left by Z shifts the rows one position down, so the rows with the Toeplitz property move one position down. At the same time, the first row is replaced with zeros, and zero at the position $1+mk$ agrees with the Toeplitz property of entries in the second row. The same analysis applies for the premultiplication by Z^l; now the Toeplitz property is translated l rows down, and zeros in the first $l-1$ rows that occurred within the Toeplitz part do agree with the Toeplitz property of other rows. Similarly, $(Z')^l A$ is Toeplitz in the domain

$$2 \le i + j \le 2n - mk. \tag{3.2}$$

Consequently, $C = BA$ coincides with a Toeplitz matrix in the intersection of the index domains (3.1) and (3.2). $\square$

Corollary 3.2. *Suppose that $A_1, \dots, A_s$ are Toeplitz banded matrices with the bandwidth $2m+1$ and that $p(x_1, \dots, x_s)$ is a homogeneous polynomial in $x_1, \dots, x_s$ of degree d. Then the matrix $A = p(A_1, \dots, A_s)$ is banded, but with $2dm+1$ diagonals, and is Toeplitz but in a smaller index domain,*

$$2 + m(d-1) \le i + j \le 2n - m(d-1).$$

Corollary 3.3. *If A is a Toeplitz banded matrix with $2m+1$ diagonals and $p_d(x)$ is a polynomial of degree d, then $B = p_d(A)$ is completely defined by the entries with indices $2 \le i+j \le 2+md$ as follows: the entries in the lower-right corner come from the persymmetric property,*

$$b_{n+1-i,n+1-j} = b_{ji}, \quad 2 \le i+j \le 2+md,$$

and all other entries come from the Toeplitz property in the remaining part of the matrix.

Moreover, let B_{1+md} denote a persymmetric matrix of order $1+md$ coinciding with $p_d(A)$ in the positions $2 \le i+j \le 2+md$. Then

$$B_{1+md} = p_d(A_{1+md}), \tag{3.3}$$

where A_{1+md} is the leading submatrix of order $1+md$ in A.

Proof. The first claim holds true because a polynomial of a persymmetric matrix remains persymmetric and because the antidiagonal $i+j = 1+md$ lies already in the Toeplitz part of B. The second claim emanates from Theorem 3.1 applied to A^k_{1+md} and to A^k with $0 \le k \le d$. These two matrices are of different orders, but we need to observe that their entries in the positions $2 \le i+j \le 2+md$ are same. ∎

These results can be generalized to the multilevel case rather straightforwardly. For convenience, let us denote by $\mathbf{m} = (m_1, \dots, m_p)$ the *half-band multiwidth*; it means that the multiwidth is $(1+2m_1, \dots, 1+2m_p)$. Also, for scalars α and β, let us agree to write

$$\alpha\mathbf{m} + \beta = (\alpha m_1 + \beta, \dots, \alpha m_p + \beta).$$

Inequalities between multi-indices are understood componentwise, and if one of the parts is a scalar α, then it must be interpreted as a multi-index $(\alpha, \dots, \alpha)$.

Theorem 3.4. *Let a p-level matrix $A = A_1 \dots A_k$ of multiorder $\mathbf{n} = (n_1, \dots, n_p)$ be a product of multi-Toeplitz (p-level Toeplitz) matrices $A_1, \dots, A_k$ of the same multiorder, and assume that each A_k is also multiband with the half-band multiwidth $\mathbf{m} = (m_1, \dots, m_p)$. Then A coincides with a p-level Toeplitz matrix everywhere in the index domain*

$$2 + (k-1)\mathbf{m} \le \mathbf{i} + \mathbf{j} \le 2\mathbf{n} - (k-1)\mathbf{m}.$$

Proof. Theorem 3.1 is evidently valid for block Toeplitz and block banded matrices. Therefore, A is a block banded matrix with blocks of order n/n_1, and these blocks in the positions $2 + (k-1)m_1 \le i_1 + j_1 \le 2n_1 - (k-1)m_1$ are part of a block Toeplitz matrix.

To examine a further structure in the blocks, we can fall back on the following general observation [23]. Let K_l denote a class of block matrices with blocks a_{ij} satisfying a homogeneous system of $\nu(K_l)$ equations

$$\gamma^s_{ij}(K_l) a_{ij} = 0, \quad 1 \le s \le \nu(K_l),$$

with given scalar coefficients $\gamma_{ij}^s(K_l)$. These coefficients and equations define the class K_l. Then let $K = K_1 \dots K_p$ denote a class of p-level matrices in which blocks of the first level satisfy the equations of K_1, blocks of the second level satisfy the equations of K_2, and so on. If σ is a permutation of indices $1, \dots, p$, then there exists a permutation matrix Π such that any matrix $A \in K_1 \dots K_p$ is transformed to a matrix $\Pi A \Pi^\top \in K_{\sigma(1)} \dots K_{\sigma(p)}$. It remains to note that, obviously,

$$\Pi A \Pi^\top = (\Pi A_1 \Pi^\top) \dots (\Pi A_p \Pi^\top),$$

and the Toeplitz and banded structures are described exactly by the equations as above. □

COROLLARY 3.5. *Suppose that $A_1, \dots, A_s$ are p-level Toeplitz multiband matrices of multiorder $\mathbf{n}$ with the half-band multiwidth $\mathbf{m}$ and that $p(x_1, \dots, x_s)$ is a homogeneous polynomial in $x_1, \dots, x_s$ of degree d. Then the matrix $A = p(A_1, \dots, A_s)$ is multiband with the half-band multiwidth $2d\mathbf{m}+1$ and simultaneously is multi-Toeplitz in the index domain*

$$2 + (d-1)\mathbf{m} \le \mathbf{i} + \mathbf{j} \le 2\mathbf{n} - (d-1)\mathbf{m}.$$

Let A be a p-level matrix of multiorder $\mathbf{n} = (n_1, \dots, n_d)$ with the entries

$$a_{\mathbf{ij}} = a(i_1, \dots, i_p\,;\, j_1, \dots, j_p).$$

A matrix A is *multilevel persymmetric* if

$$a(i_1, \dots, n_l + 1 - i_l, \dots, i_p \;;\; j_1, \dots, n_l + 1 - j_l, \dots, j_p) =$$

$$a(i_1, \dots, j_l, \dots, i_p \;;\; j_1, \dots, i_l, \dots, j_p),$$

$$1 \le i_l \le n_l, \quad 1 \le l \le p.$$

As is readily seen, A is completely recovered from the entries in the index domain

$$2 \le \mathbf{i} + \mathbf{j} \le 1 + \mathbf{n}.$$

If $\mathbf{h}$ is a multi-index from the range $1 \le \mathbf{h} \le \mathbf{n}$, then $A_{\mathbf{h}}$ will stand for a *multilevel leading* submatrix in A, defined as a p-level matrix $\hat{A} = [\hat{a}_{\mathbf{ij}}]$ of multiorder $\mathbf{h}$ with the entries

$$\hat{a}_{\mathbf{ij}} = a_{\mathbf{ij}}, \quad 1 \le \mathbf{i}, \mathbf{j} \le \mathbf{h}.$$

COROLLARY 3.6. *Assume that A is a multi-Toeplitz multiband matrix of multiorder $\mathbf{n}$ with half-band multiwidth $2\mathbf{m}+1$ diagonals, and let $p_d(x)$ be a polynomial of degree d. Then $B = p_d(A)$ is completely defined by the entries with multi-indices*

$$2 \le \mathbf{i} + \mathbf{j} \le 2 + d\mathbf{m}$$

by the muiltilevel persymmetric property in the block corners and by the multilevel Toeplitz property in the remaining part of the matrix.

Moreover, let $B_{1+d\mathbf{m}}$ denote a multilevel persymmetric matrix of order $1 + md$ which is a multilvel leading submatrix of multiorder $1+d\mathbf{m}$ in $p_d(A)$, e.g., it coincides with $p_d(A)$ in the positions $1 \le i + j \le 1 + md$. Then

$$B_{1+d\mathbf{m}} = p_d(A_{1+d\mathbf{m}}),$$

where $A_{1+d\mathbf{m}}$ *is the multilevel leading submatrix in* A.

An obvious and useful consequence is a simple $O(1)$ algorithm for computing $p_d(A)$, where A is a multi-Toeplitz multiband matrix. All we need to do is the computation of $p_d(A_{1+d\mathbf{m}})$. The cost depends on d and $\mathbf{m}$ but is not affected by $\mathbf{n}$.

It is also useful to remark that the numerical range of a multi-Toeplitz matrix is easy to find, especially in the multiband case. To this end, a matrix $A = [a_{\mathbf{kl}}]$ is associated with a formal p-variate Fourier series

$$f(\mathbf{x}) = \sum_{\mathbf{k}} a_{\mathbf{k}} \exp(\mathbf{ikx}), \quad \mathbf{i} = \sqrt{-1}, \tag{3.4}$$

$$\mathbf{k} = (k_1, \ldots, k_p), \quad \mathbf{x} = (x_1, \ldots, x_p), \quad \mathbf{kx} = k_1x_1 + \cdots + k_px_p.$$

If A is multiband, then the series has finitely many terms and obviously defines a trigonometric polynomial with the coefficients

$$a_{\mathbf{k}} = \frac{1}{(2\pi)^p} \int\limits_{[0,2\pi]^p} f(\mathbf{x}) \exp(-\mathbf{ikx}) d\mathbf{x}. \tag{3.5}$$

More generally, let $f(\mathbf{x})$ be an arbitrary p-variate function which is 2π-periodic in each variable and is such that all integrals in the right-hand side of (3.5) exist. Then the series associated with f is the right-hand side of (3.4) with $a_{\mathbf{k}}$ defined by (3.5). Then the Fourier coefficients $a_{\mathbf{k}}$ give rise to a family of multi-Toeplitz matrices

$$A_{\mathbf{n}} = [a_{\mathbf{k}-\mathbf{l}}], \quad 1 \le \mathbf{k}, \mathbf{l} \le \mathbf{n}.$$

In regard to this family, f is referred to as a *generating function* or *symbol*. The next theorem seems to be well known in the univariate case, but it holds true as well for multi-Toeplitz matrices.

THEOREM 3.7. *The numerical range of a multi-Toeplitz matrix* A *with a continuous generating function* $f(\mathbf{x})$ *belongs to the closure of the convex hull of all the values of* $f(\mathbf{x})$.

Proof. Consider a vector $v = [v_l]$ as a multilevel vector of multisize $\mathbf{n}$; this means that we write $v_l = v_{\mathbf{l}}$ with the natural one-to-one correspondence

$$l \leftrightarrow \mathbf{l} = (l_1, \ldots, l_p).$$

Then introduce a trigonometric polynomial

$$v(\mathbf{x}) = \sum_{0 \le \mathbf{l} \le \mathbf{n}} v_{\mathbf{l}} \exp(\mathbf{ilx}),$$

and observe that

$$(A_{\mathbf{n}}v, v) = \frac{1}{(2\pi)^p} \int\limits_{[0,2\pi]^p} f(\mathbf{x}) \left|v(\mathbf{x})\right|^2 d\mathbf{x}.$$

Since the identity is a p-level Toeplitz matrix with the generating function $f(x) \equiv 1$, it follows that

$$(v, v) = \frac{1}{(2\pi)^p} \int\limits_{[0,2\pi]^p} \left|v(\mathbf{x})\right|^2 d\mathbf{x}.$$

It remains to note that the integrals can be approximated by quadrature sums on uniform grids, and, consequently, the ratio $(A_{\mathbf{n}}v, v)/(v, v)$ is a limit of convex combinations of the values of $f(\mathbf{x})$ on uniform grids. ∎

4. Numerical examples.

4.1. Experiment 1. Consider a multiband multi-Toeplitz matrix A with two levels of multiorder $\mathbf{n} = (n, n)$. In the two-level tridiagonal case, A is defined by the multilevel leading submatrix of multiorder $\mathbf{m} = (2, 2)$. Consider a particular case as follows:

$$A_{\mathbf{m}} = \left[\begin{array}{cc|cc} a_0 & a_{-1} & a_{-1} & 0 \\ a_1 & a_0 & 0 & a_{-1} \\ \hline a_1 & 0 & a_0 & a_{-1} \\ 0 & a_1 & a_1 & a_0 \end{array}\right],$$

where

$$a_0 = 4s + c, \quad a_{-1} = -s(1+h), \quad a_1 = -s(1-h).$$

The generating function for A is

$$\begin{aligned} f(x_1, x_2) &= 4s + c - s(1+h)\exp(-\mathbf{i}x_1) - s(1-h)\exp(\mathbf{i}x_1) \\ &\quad -s(1+h)\exp(-\mathbf{i}x_2) - s(1-h)\exp(\mathbf{i}x_2) \\ &= 4 - 2s\cos x_1 - 2s\cos x_2 - \mathbf{i}\left(2sh\sin x_1 + 2sh\sin x_2\right). \end{aligned}$$

Set

$$s = 1/4, \quad c = 3.$$

Then the numerical range of A belongs to the rectangle $[3, 5] \times [-h, h]$. The circumscribed ellipse can be obtained by an affine mapping of the Bernstein ellipse with the parameter

$$r' = \sqrt{\frac{1+h}{1-h}}.$$

The corresponding affine mapping is

$$z = \xi + \eta\, w, \quad \xi = 4, \quad \eta = \frac{2}{r' + 1/r'}.$$

Let us take a larger cofocal ellipse (an analyticity domain for functions to be considered) as the one passing through $z = 0$, with the related Bernstein ellipse defined by the parameter

$$r = \frac{\xi + \sqrt{\xi^2 - \eta^2}}{\eta}.$$

According to (2.4), the decay rate (accuracy of polynomial approximations) is determined by $\rho = r'/r$. The approximation error will be measured in the infinity norm

$$\mathtt{ERROR} = ||f(A) - p_d(A)||_\infty / ||f(A)||_\infty.$$

Below d is the degree of the interpolation polynomial, and N is the matrix size. In Table 1 we present the errors of approximation to $f(A) = A^{-1}$ by polynomials of

TABLE 1
Relative error for $f(A) = A^{-1}$ versus h, $d = 7$, $N = 900$.

h	0.1	0.3	0.5	0.7	0.9
ERROR	$7 \cdot 10^{-7}$	$1 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$6 \cdot 10^{-6}$	$1 \cdot 10^{-5}$
$(r'/r)^d$	$1 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$9 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$4 \cdot 10^{-5}$

TABLE 2
Relative errors for polynomial approximations, $h = 1/2$, $N = 900$.

d	3	5	7	9	11	13
A^{-1}	$2 \cdot 10^{-3}$	$7 \cdot 10^{-5}$	$3 \cdot 10^{-6}$	$1 \cdot 10^{-7}$	$4 \cdot 10^{-9}$	$1 \cdot 10^{-10}$
$\exp(A)$	$1 \cdot 10^{-2}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-6}$	$1 \cdot 10^{-8}$	$5 \cdot 10^{-11}$	$8 \cdot 10^{-13}$
$\log(A)$	$4 \cdot 10^{-4}$	$1 \cdot 10^{-5}$	$3 \cdot 10^{-7}$	$8 \cdot 10^{-9}$	$2 \cdot 10^{-10}$	$8 \cdot 10^{-12}$

degree $d = 7$ in dependence on h. In Table 2 we expose the errors for basic matrix functions in dependence on the degree of polynomial.

Note that the polynomial approximations are computed by the $O(1)$ algorithm of section 4. It actually finds a polynomial of a small-size multilevel leading submatrix of A. Then we easily get all the entries of $f(A)$ without further computations. The computation time does not depend on the matrix size $N = n^2$ or the accuracy, the latter being confirmed in Table 3. The reference computation is made by a call to the MATLAB function `expm`, and its running time grows as N^3; consequently, this standard method is not feasible for large N. In contrast, our algorithm will compute this matrix function in the same time for arbitrarily large N.

TABLE 3
Relative errors versus matrix size, $f(A) = \exp(A)$, $d = 9$, $h = 1/2$.

$N = n^2$	400	900	1600	2500
ERROR	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$

4.2. Experiment 2. For another example, take $q = 0.1$, and consider two-level matrices with the entries

$$a_{i_1 i_2; j_1 l_2} = \begin{cases} \exp\{-q((i_1 - j_1)^2 + (i_2 - j_2)^2)\}, & i \neq j, \\ 4, & i = j. \end{cases}$$

Take the level sizes to be $n_1 = n_2 = 30$, and examine the decay properties at each level for basic matrix functions of A. Tables 4–6 show the actual decay results of two-level diagonals $i_1 - j_1 = k_1,\ k_2 = i_2 - j_2$ as functions of k_1, k_2. The rate of decay agrees with our theoretical estimates.

The decay properties can serve as a base for some fast algorithms rather than the polynomial approximations considered above. A general scheme is proposed in [15], developing an approach considered for the approximate matrix inversion using Toeplitz-like structures in [20] and using tensor approximations in [19].

Given a two-level matrix A with the level bandwidths m_1, m_2, in order to perform the inversion, we can consider the approximate Newton–Schultz iteration as follows:

$$X_{k+1} = \mathcal{P}_{k_1,k_2}(X_k - X_k A X_k),$$

where $\mathcal{P}_{k_1,k_2}$ is a truncation operator that sets to zero all the entries outside the two-level band with bandwidth parameters k_1, k_2.

TABLE 4
Decay in two-level diagonals: $f(A) = A^{-1}$, $n_1 = n_2 = 30$.

k_1, k_2	0	1	3	5	7	9	11	12
0	9.E+00	7.E-01	1.E-01	3.E-02	5.E-03	1.E-03	7.E-04	1.E-04
1	7.E-01	5.E-01	6.E-02	3.E-02	5.E-03	9.E-04	6.E-04	8.E-05
3	1.E-01	6.E-02	4.E-02	1.E-02	6.E-03	4.E-04	4.E-04	9.E-05
5	3.E-02	3.E-02	1.E-02	5.E-03	2.E-03	1.E-03	4.E-05	2.E-04
7	5.E-03	5.E-03	6.E-03	2.E-03	9.E-04	4.E-04	2.E-04	1.E-04
9	1.E-03	9.E-04	4.E-04	1.E-03	4.E-04	2.E-04	6.E-05	8.E-06
11	7.E-04	6.E-04	4.E-04	4.E-05	2.E-04	6.E-05	3.E-05	3.E-05
12	1.E-04	8.E-05	9.E-05	2.E-04	1.E-04	8.E-06	3.E-05	1.E-05

TABLE 5
Decay in two-level diagonals: $f(A) = \log(A)$, $n_1 = n_2 = 30$.

k_1, k_2	0	1	3	5	7	9	11	12
0	4.E+01	4.E+00	1.E+00	5.E-02	1.E-02	5.E-03	9.E-04	6.E-04
1	4.E+00	3.E+00	8.E-01	6.E-02	1.E-02	5.E-03	9.E-04	5.E-04
3	1.E+00	8.E-01	9.E-02	6.E-02	3.E-03	3.E-03	9.E-04	3.E-04
5	5.E-02	6.E-02	6.E-02	9.E-03	7.E-03	2.E-04	5.E-04	4.E-05
7	1.E-02	1.E-02	3.E-03	7.E-03	1.E-03	9.E-04	2.E-05	1.E-04
9	5.E-03	5.E-03	3.E-03	2.E-04	9.E-04	1.E-04	1.E-04	6.E-05
11	9.E-04	9.E-04	9.E-04	5.E-04	2.E-05	1.E-04	2.E-05	1.E-05
12	6.E-04	5.E-04	3.E-04	4.E-05	1.E-04	6.E-05	1.E-05	2.E-05

TABLE 6
Decay in two-level diagonals: $f(A) = \exp\{-A\}$, $n_1 = n_2 = 30$.

k_1, k_2	0	1	3	5	7	9	11	12
0	1.E+00	2.E-01	8.E-03	8.E-03	4.E-03	1.E-03	2.E-04	4.E-04
1	2.E-01	1.E-01	7.E-03	7.E-03	4.E-03	1.E-03	2.E-04	4.E-04
3	8.E-03	7.E-03	2.E-02	2.E-03	2.E-03	1.E-03	2.E-04	4.E-04
5	8.E-03	7.E-03	2.E-03	4.E-03	8.E-04	7.E-04	4.E-04	2.E-04
7	4.E-03	4.E-03	2.E-03	8.E-04	1.E-03	2.E-04	2.E-04	3.E-05
9	1.E-03	1.E-03	1.E-03	7.E-04	2.E-04	3.E-04	5.E-05	1.E-04
11	2.E-04	2.E-04	2.E-04	4.E-04	2.E-04	5.E-05	9.E-05	4.E-05
12	4.E-04	4.E-04	4.E-04	2.E-04	3.E-05	1.E-04	4.E-05	3.E-06

TABLE 7
Residual values for the Newton–Schultz iteration, $m = 7$, $k = 21$.

iteration	1	2	3	4	5	6	7
$n = 50$	3.4E+01	2.5E+01	1.3E+01	3.4E+00	2.5E-01	1.4E-03	1.2E-04
$n = 100$	6.9E+01	4.9E+01	2.5E+01	6.8E+00	5.0E-01	2.7E-03	2.8E-04

Consider the above matrix with level sizes $n = n_1 = n_2$. Let the input bandwidth be $m = m_1 = m_2$, and set the truncation bandwidth parameter to $k = k_1 = k_2$. The initial guess X_0 is chosen as $X_0 = \alpha I$; below we set $\alpha = 0.05$. Table 7 shows the residual values $r_i = ||I - X_k A||_F$.

5. Conclusions. In this paper we presented a simple framework that allows us to revisit a proof of the Benzi–Golub result [3] on decay properties for matrix functions of a banded Hermitian matrix, and we also extend this result to multilevel multibanded matrices. However, despite the generality of our theorems, we ought to note that practical usefulness of the presented decay estimates depends on the involved constants.

Some numerical examples are given to illustrate the multilevel decay properties on typical matrix functions such as the inverse, exponential, and logarithm. We show how the decay properties can be used in approximate computation of the inverse matrix by the Newton–Schultz iteration with a sparsification on every step so that the total number of nonzero entries is kept limited.

We also propose $O(1)$ algorithms for approximating matrix functions of multiband multi-Toeplitz matrices.

REFERENCES

[1] Z. Bai, M. Fahey, and G. H. Golub, *Some large-scale matrix computation problems*, J. Comput. Appl. Math., 74 (1996), pp. 71–89.

[2] B. Beckermann, S. A. Goreinov, and E. E. Tyrtyshnikov, *Some remarks on the Elman estimate for GMRES*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 772–778.

[3] M. Benzi and G. H. Golub, *Bounds for the entries of matrix fucntions with application to preconditioning*, BIT, 39 (1999), pp. 417–438.

[4] M. Benzi and N. Razouk, *Decay bounds and $O(n)$ algorithms for approximating functions of sparse matrices*, Electron. Trans. Numer. Anal., 28 (2007), pp. 16–39.

[5] L. Bergamaschi, M. Caliari, and M. Vianello, *Efficient approximation of the exponential operator for discrete 2D advection-diffusion problems*, Numer. Linear Algebra Appl., 10 (2003), pp. 271–289.

[6] I. A. Blatov, *On algebras and applications of operators with pseudosparse matrices*, Sibirsk. Mat. Zh., 37 (1996), pp. 36–59.

[7] W. Ching and M. Ng, *Markov Chains: Models, Algorithms and Applications*, Internat. Ser. Oper. Res. Management Sci. 83, Springer, New York, 2006.

[8] M. Crouzeix, *Numerical Range and Functional Calculas in Hilbert Space*, J. Funct. Anal., 244 (2007), pp. 668–690.

[9] B. Delyon and F. Delyon, *Generalization of Von Neumann spectral sets and integral representation of operators*, Bull. Soc. Math. France, 1 (1999), pp. 25–42.

[10] S. Demko, *Inverses of band matrices and local convergence of spline projections*, SIAM J. Numer. Anal., 14 (1977), pp. 616–619.

[11] S. Demko, W. F. Moss, and P.W. Smith, *Decay rates for inverses of band matrices*, Math. Comp., 43 (1984), pp. 491–499.

[12] V. L. Druskin and L. A. Knizhnerman, *The polynomial methods of calculating functions of symmetric matrices*, USSR Comput. Maths. Math. Phys., 29 (1980), pp. 112–121.

[13] A. O. Gel'fand, *Calculus of Finite Differences*, Hidustan Publishing, Delhi, India, 1971.

[14] J. Gutierrez-Gutierrez, P. Crespo, and A. Boettcher, *Functions of banded Hermitian block Toeplitz matrices in signal processing*, Linear Algebra Appl., 422 (2007), pp. 788–807.

[15] W. Hackbusch, B. N. Khoromskij, and E. E. Tyrtyshnikov, *Approximate iterations for structured matrices*, Numer. Math., 109 (2008), pp. 365–383.

[16] N. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.

[17] M. Ng, G. Qiu, and A. Yip, *A study of interactive multiple class image segmentation problems*, Internat. J. Imaging Syst. Technol., 20 (2010), pp. 191–201.

[18] V. Olshevsky, I. Oseledets, and E. Tyrtyshnikov, *Tensor properties of multilevel Toeplitz and related matrices*, Linear Algebra Appl., 412 (2006), pp. 1–21.

[19] I. V. Oseledets and E. E. Tyrtyshnikov, *Approximate inversion of matrices in the process of solving a hypersingular integral equation*, Comput. Math. Math. Phys. 45 (2005), pp. 302–313.

[20] V. Y. Pan and Y. Rami, *Newton's iteration for the inversion of structured matrices*, in Structured Matrices: Recent Developments in Theory and Computation, Advances in Computation, D. A. Bini, E. Tyrtyshnikov, and P. Yalamov, eds., Nova Science Publishers, Huntington, New York, 2001, pp. 79–90.

[21] Q. Yang, Z. Huang, and M. Ng, *A data cube model for prediction-based Web prefetching*, J. Intell. Inf. Syst., 20 (2003), pp. 11–30.

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

[23] V. V. Voevodin and E. E. Tyrtyshnikov, *Computational Processes with Toeplitz Matrices*, Nauka, Moscow, 1987 (in Russian).