Approximate Iterations for Structured Matrices

Wolfgang Hackbusch, Boris N. Khoromskij
Max-Planck-Institut für Mathematik in den Naturwissenschaften,
Inselstr. 22-26, D-04103 Leipzig, Germany; {wh, bhk}@mis.mpg.de

Eugene E. Tyrtyshnikov *
Institute of Numerical Mathematics, Russian Academy of Sciences,
Gubkina 8, 119991 Moscow, Russia; tee@inm.ras.ru

Abstract

Important matrix-valued functions $f(A)$ are, e.g., the inverse A^{-1}, the square root $\sqrt{A}$ and the sign function. Their evaluation for large matrices arising from pdes is not an easy task and needs techniques exploiting appropriate structures of the matrices A and $f(A)$ (often $f(A)$ possesses this structure only approximately). However, intermediate matrices arising during the evaluation may lose the structure of the initial matrix. This would make the computations inefficient and even infeasible. However, the main result of this paper is that an iterative fixed-point like process for the evaluation of $f(A)$ can be transformed, under certain general assumptions, into another process which preserves the convergence rate and benefits from the underlying structure. It is shown how this result applies to matrices in a tensor format with a bounded tensor rank and to the structure of the hierarchical matrix technique. We demonstrate our results by verifying all requirements in the case of the iterative computation of A^{-1} and $\sqrt{A}$. The exact iteration is analysed in the case of sign(A), here, however, the iteration is constrained to a subspace and does not satisfy the assumptions of our theorems.

AMS Subject Classification: 65F30, 65F50, 65N35, 65F10
Key words: iterative algorithms, structured matrices, matrix functions, matrix approximations, low-rank matrices, hierarchical matrices, Kronecker products, multi-dimensional matrices, tensors, integral operators, numerical methods

1 Introduction

We consider important matrix-valued functions $f(A)$ as, e.g., the inverse A^{-1}, the square root $\sqrt{A}$ and the sign function. In particular, we are interested in evaluations of $f(A)$ for

*This work was performed during the stay of the third author at the Max-Planck-Institute for Mathematics in the Sciences (Leipzig) and also supported by the Russian Fund of Basic Research (grants 05-01-00721, 04-07-90336) and a Priority Research Grant of the Department of Mathematical Sciences of the Russian Academy of Sciences.

matrices A arising from partial differential equations. Obviously, the computation of $f(A)$ for large-scale matrices A is not an easy task. In the numerical treatment one has to avoid the full-matrix representation. Instead one should use special representations (i.e., special data structures) which, on the other hand, correspond to special properties of the argument A, of the result $f(A)$ and of the auxiliary matrices arising during the computation process.

Examples for such a representation are Toeplitz-like structures or a sparse-matrix format. The latter format is not successful for our examples, since sparse matrices A produce results A^{-1}, $\sqrt{A}$, $\operatorname{sign}(A)$, which are usually non-sparse and which, moreover, cannot be approximated by sparse matrices. This is different for the format of hierarchical matrices (cf. [18, 16, 19, 20]) and the hierarchical Kronecker-tensor product (HKT) representation (cf. [24, 21, 22]).

The matrices belonging to a particular representation are characterised by a subset S of the vector space of matrices. The letter S abbreviates "structured matrices". In the simplest case, $A \in S$ implies $f(A) \in S$. If also all intermediate results belong to S, the whole computational process can be performed using the special data structures of S. The purpose of this paper is the analysis of a more complicated situation, when $A \in S$ does not imply $f(A) \in S$, but $f(A)$ has a good approximation in S. We illustrate this situation by the following example.

We consider a discrete two-dimensional Laplacian

$$A = T \otimes I + I \otimes T, \quad T = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}, \tag{1.1}$$

where T and the identity I are $n \times n$ matrices. Obviously, A is a matrix of size $n^2 \times n^2$ with a very special structure: it is *exactly* the sum of two terms, each being the Kronecker (tensor) product of two $n \times n$ matrices. It is remarkable that A^{-1} is *approximately* of the same structure but with a greater number of terms. This number is called the tensor rank; the way the rank depends on the approximation accuracy ε and n can be seen from Table 1.1:

ε n	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
20	4	5	6	7	8	9	10	10
40	4	6	7	8	10	11	12	13
80	4	6	8	10	11	13	14	15
160	4	7	9	11	13	14	16	18
320	5	7	10	12	14	16	18	20

Table 1.1: Tensor ranks for ε-approximations to A^{-1}.

We observe a logarithmic growth of the tensor rank upon ε and as well upon n. More precisely, the rank estimate $r = O(|\log \varepsilon| \log n)$ can be proven (cf. [23]) based on approximation by exponential sums also for Kronecker products involving more than two factors (cf. [15, 21, 22]). Thus, A^{-1} can be approximated by a matrix defined by a reasonably small number of parameters in the tensor format.

So far the existence of an approximation $\tilde{B} \approx A^{-1}$ with $\tilde{B} \in S$ is ensured (here, the set S of structured matrices is given by sums of Kronecker products with a certain limited number of terms). It remains to design an algorithm for computing $f(A) = A^{-1}$. A possible choice is the Newton-Schulz iteration

$$X_0 = \alpha I, \qquad X_k = X_{k-1}(2I - AX_{k-1}) \qquad\qquad (k = 1, 2, \ldots).$$

For this iteration it can be proved that if $0 < \alpha \leq 1/4$ then $X_k \to A^{-1}$, and the convergence is quadratic.

Here the important question arises, whether the *intermediate matrices* X_k belong to the subset S or can be well approximated by $\tilde{X}_k \in S$. In the following numerical experiment each X_k admits a suitable approximation of low tensor rank as can be seen from Table 1.2:

k ε	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
10^{-3}	2	3	4	4	5	5	6	6	5	6	6	6	7	7	7	7
10^{-6}	2	4	7	8	8	9	10	10	11	12	12	13	14	14	13	13

Table 1.2: Tensor ranks for ε-approximations to X_k ($n = 160$).

Thus, a natural idea is to substitute X_k by its approximation in the tensor format. Such a substitution is called *truncation.* Assume that the truncation is performed at every iteration. Then the following questions arise: How will this affect the convergence rate of the Newton-Schulz method? Will the convergence remain quadratic? The answers are positive. Moreover, the same answer is valid not only for the Laplacian but typical for the truncation based on the tensor or hierarchical formats [17, 27, 21, 22].

The main result of this paper is that an iterative fixed-point process for the evaluation of $f(A)$ can be transformed, under certain general assumptions, into another process which preserves the convergence rate and benefits from the underlying structure. It is shown how this result applies to matrices in a tensor format with a bounded tensor rank and to the structure of the hierarchical matrix technique. We demonstrate our results by verifying all requirements in the case of the iterative computation of A^{-1} and $\sqrt{A}$.

In this paper we propose a general framework in which the above-mentioned results appear as particular cases. Our main results are two theorems (Section 2) that turn out to be both entirely general and quite elementary. Despite the latter, they do not seem to be well-known in the community of numerical analysis and structured matrices. Our results clearly amplify the role of nonlinear iterative schemes in computations with structured matrices. It is especially gainful that they apply to many interesting iterative scheme and various classes of structured matrices including those already in work and those that may appear yet in application contexts.

Nevertheless, there are iterations which do not satisfy the requirements of our theorems. For instance, our theory does not apply when the success of the iteration depends on the fact that the iterates X_k stay in some sub-manifold. We give examples of such “constrained” iterations for computing $\sqrt{A}$ and $\operatorname{sign}(A)$ in §4.2.2 and §4.3 and analyse the convergence of the exact iteration. Although the truncated iteration is not supported by our theorems, the numerical performance is reasonable.

The rest of the paper is organised as follows.

In Section 2 we consider an iteration $X_k = \Phi_k(X_{k-1})$, which starting with $X_0 := \Phi_0(A)$ is assumed to converge to $f(A)$. The quadratic convergence is described in detail in Lemma 2.1. Next we introduce a so-called *truncation operator* R which maps into a subset S (which we call the set of "structured" elements). The combination of the iteration with the truncation operator yields the truncated iteration $Y_k = R(\Phi_k(Y_{k-1}))$. In Theorem 2.2 we describe the characteristic requirements on R so that the truncated iteration has similar convergence properties as the original iteration. The final Theorem 2.4 considers the important case that the desired result $f(A)$ does not belong to S but is close to S.

It remains to verify that the assumptions on the truncation operator R can be satisfied in practically relevant cases. In Section 3 we describe a general framework which is later applied (i) to the structure used in the hierarchical matrix technique and (ii) to low Kronecker rank matrices.

Section 4 is devoted to the convergence analysis of certain matrix iterations resulting in A^{-1}, $\sqrt{A}$ and $\operatorname{sign}(A)$. In particular, the general theory from Sections 2 and 3 ensures the quadratic convergence of the truncated non-constrained iterations to compute A^{-1} and $\sqrt{A}$. In §4.2.2 we describe an example of an iteration for $\sqrt{A}$ which is constrained to a subspace. An iteration of the same type for $\operatorname{sign}(A)$ is given in §4.3.

2 Main result

2.1 Exact iteration

Let V be a normed space V and consider a function $f : V \to V$ and $A \in V$. Assume that $B := f(A)$ can be obtained by an iteration of the form

$$X_k = \Phi_k(X_{k-1}), \qquad k = 1, 2, \ldots , \tag{2.1}$$

where Φ_k is a one-step operator. Further, assume that for any initial guess X_0 sufficiently close to B, the process converges:

$$\lim_{k\to\infty} X_k = B. \tag{2.2}$$

If $\Phi_k = \Phi$ does not depend on k, (2.1) represents the important *fixed-point iteration.*

Lemma 2.1 *Let B and Φ_k be as above and assume that there are constants c_Φ, $\varepsilon_\Phi > 0$ and $\alpha > 1$ such that*

$$\|\Phi_k(X) - B\| \;\leq\; c_\Phi \,\|X - B\|^\alpha \quad \text{for all } X \in V \text{ with } \|X - B\| \leq \varepsilon_\Phi \text{ and all } k \in \mathbb{N}, \tag{2.3}$$

and set

$$\varepsilon := \min\left(\varepsilon_\Phi,\, 1/c\right), \quad c := \sqrt[\alpha-1]{c_\Phi}\,. \tag{2.4}$$

Then (2.2) holds for any initial guess X_0 satisfying $\|X_0 - B\| < \varepsilon$, and, moreover,

$$\|X_k - B\| \;\leq\; c^{-1} \left(c\; \|X_0 - B\|\,\right)^{\alpha^k} \qquad (k = 0, 1, 2, \ldots)\;. \tag{2.5}$$

Proof. Let $e_k := \|X_k - B\|$. Then, due to (2.3),

$$e_k \;\le\; c_\Phi e_{k-1}^\alpha, \qquad \text{provided that } e_{k-1} \le \varepsilon_\Phi\,. \tag{2.6}$$

Because of (2.6), the inequalities $e_{k-1} \le \varepsilon \le \varepsilon_\Phi$ imply $e_k \le c_\Phi \varepsilon^\alpha = c^{\alpha-1}\varepsilon^\alpha = \varepsilon\,(c\varepsilon)^{\alpha-1} \le \varepsilon$. Hence, all iterates stay in the ε-neighbourhood of B. (2.5) is proved by induction:

$$e_k \underset{(2.6)}{\le} c_\Phi e_{k-1}^\alpha \underset{\text{induction hypothesis}}{=} c_\Phi \cdot \left(c^{-1}\,(ce_0)^{\alpha^{k-1}}\right)^\alpha \underset{c_\Phi = c^{\alpha-1}}{=} c^{\alpha-1}\cdot c^{-\alpha}\,(ce_0)^{\alpha^k} = c^{-1}\,(ce_0)^{\alpha^k}\,.$$

Whenever $e_0 < \varepsilon$, (2.5) shows $e_k \to 0$. ∎

We remark that (2.6) together with $e_0 \le \varepsilon$ implies monotonicity:

$$\|X_k - B\| \;\le\; \|X_{k-1} - B\|\,. \tag{2.7}$$

2.2 Truncated iteration

Let $S \subset V$ be a subset (not necessarily a subspace) considered as a class of certain structured elements (e.g., matrices of a certain data structure) and suppose that $R : V \to S$ is an operator from V onto S. We call R a *truncation operator*. It is assumed that $R(X) = X$ for any $X \in S$ (i.e., all elements in S are fixed points of R). Note that, in general, R is a nonlinear mapping. The truncation of real numbers to machine numbers is a common example for $V = \mathbb{R}$.

Now, instead of (2.1), consider a *truncated iterative process* defined as follows:

$$\begin{aligned} Y_0 &:= R(X_0), \\ Y_k &:= R(\Phi_k(Y_{k-1})) \qquad (k = 1,\, 2\ldots.)\,. \end{aligned} \tag{2.8}$$

The next theorem needs the assumption that the desired result $B := f(A)$ belongs (exactly) to the subset S. Later, in Theorem 2.4, this requirement will be relaxed.

Theorem 2.2 *Under the premises of Lemma 2.1, assume that*

$$\|X - R(X)\| \;\le\; c_R\,\|X - B\| \qquad \textit{for all } X \in V \textit{ with } \|X - B\| \le \varepsilon_\Phi\,. \tag{2.9}$$

Then there exists $\delta > 0$ such that the truncated iterative process (2.8) converges to B so that

$$\|Y_k - B\| \;\le\; c_{R\Phi}\,\|Y_{k-1} - B\|^\alpha \quad \textit{with } c_{R\Phi} := (c_R + 1)c_\Phi \qquad (k = 1,\, 2, \ldots) \tag{2.10}$$

for any starting value $Y_0 = R(Y_0)$ satisfying $\|Y_0 - B\| < \delta$.

Proof. Let ε as in (2.4) and define $Z_k := \Phi_k(Y_{k-1})$. By (2.7) we have $\|Z_k - B\| \le \|Y_{k-1} - B\|$, provided that $\|Y_{k-1} - B\| \le \varepsilon$. Then

$$\|Y_k - B\| = \|(R(Z_k) - Z_k) + (Z_k - B)\| \le (c_R + 1)\,\|Z_k - B\|\,. \tag{2.11a}$$

Assuming $\|Y_{k-1} - B\| \le \varepsilon$, the inequalities $\varepsilon \le \varepsilon_\Phi$ and (2.3) ensure

$$\|Z_k - B\| = \|\Phi_k(Y_{k-1}) - B\| \le c_\Phi\,\|Y_{k-1} - B\|^\alpha\,. \tag{2.11b}$$

Combining (2.11a) and (2.11b), we obtain (2.10) for any k, provided that $\|Y_{k-1} - B\| \le \varepsilon$.

Similar to the proof of Lemma 2.1 and (2.7), the choice

$$\delta := \min\left(\varepsilon,\, 1/C\right), \qquad C := \sqrt[\alpha-1]{c_{R\Phi}} \tag{2.11c}$$

guarantees that $\|Y_0 - B\| \le \delta$ implies $\|Y_k - B\| \le \delta \le \varepsilon$ for all $k \in \mathbb{N}$. ∎

Corollary 2.3 *Under the assumptions of Theorem 2.2, any starting value Y_0 with $\|Y_0 - B\| \le \delta$ leads to*

$$\|Y_k - B\| \;\le\; C^{-1}\left(C\,\|Y_0 - B\|\right)^{\alpha^k} \qquad (k = 1, 2, \ldots), \tag{2.12}$$

where C and δ are defined in (2.11c).

2.3 The case of $B \notin S$

In most of the practical applications, the desired result B will not belong to the subset S, but may be close to S. The following requirement (2.14) states that $\|B - R(B)\| \le \varepsilon_{RB}$. Then the truncated iteration cannot converge to B, but it comes sufficiently close to B. In fact, in a first phase the truncated iteration is described by (2.12) with C replaced by $C' := \sqrt[\alpha-1]{2c_{R\Phi}}$ until it reaches the $2\varepsilon_{RB}$-neighbourhood of B. The quantity ε_{RB} must be sufficiently small:

$$\varepsilon_{RB} < \frac{\eta}{2}, \qquad \text{where } \eta := \min\left(\varepsilon, 1/\sqrt[\alpha-1]{2c_{R\Phi}}\right) \tag{2.13}$$

with $c_{R\Phi} = (c_R + 1)c_\Phi$ as defined above.

Theorem 2.4 *Under the premises of Lemma 2.1, suppose*

$$\|X - R(X)\| \;\le\; c_R\,\|X - B\| + \varepsilon_{RB} \qquad \textit{for all } X \in V \textit{ with } \|X - B\| \le \varepsilon_\Phi, \tag{2.14}$$

where ε_{RB} satisfies (2.13). Further, assume $\|Y_0 - B\| < \eta$ and define Y_k by the truncated iteration (2.8). Let m be the minimal $k \in \mathbb{N}$ such that

$$\|Y_{k-1} - B\|^\alpha \;\le\; \frac{\varepsilon_{RB}}{c_{R\Phi}}. \tag{2.15}$$

Then the errors $\|Y_k - B\|$ strictly decrease for $1 \le k < m$, while for $k \ge m$ the iterates stagnate in a $2\varepsilon_{RB}$-neighbourhood of the true result:

$$\|Y_k - B\| \;\le\; \begin{cases} 2c_{R\Phi}\,\|Y_{k-1} - B\|^\alpha & \textit{for } k \le m-1, \\ 2\varepsilon_{RB} & \textit{for } k \ge m. \end{cases} \tag{2.16}$$

Proof. Instead of (2.11a) we now have

$$\|Y_k - B\| \le \|Y_k - Z_k\| + \|Z_k - B\| \le (c_R + 1)\,\|Z_k - B\| + \varepsilon_{RB},$$

which obviously implies

$$\|Y_k - B\| \;\le\; c_{R\Phi}\,\|Y_{k-1} - B\|^\alpha + \varepsilon_{RB}. \tag{2.17}$$

If $k < m$, the inequality $\varepsilon_{RB} \le c_{R\Phi}\,\|Y_{k-1} - B\|^\alpha$ holds and implies $\|Y_k - B\| \le 2c_{R\Phi}\,\|Y_{k-1} - B\|^\alpha$. Hence, (2.10) holds with $c_{R\Phi}$ replaced by $2c_{R\Phi}$ giving rise to (2.12) with C replaced by $C' := \sqrt[\alpha-1]{2c_{R\Phi}}$. The initial error estimate $\|Y_0 - B\| < \eta$ implies the strict decrease of $\|Y_k - B\|$ until (2.15) holds.

If $k = m$, (2.17) shows $\|Y_m - B\| \le 2\varepsilon_{RB}$. For $k \ge m$, the estimate $c_{R\Phi}\,(2\varepsilon_{RB})^\alpha + \varepsilon_{RB} \le 2\varepsilon_{RB}$ derived from (2.13) proves the second case in (2.16). ■

Corollary 2.5 *Theorems 2.2 and 2.4 can be generalised by replacing the conditions (2.9) and (2.14) with the respective inequalities*

$$\|(I-R)(X)\| \;\leq\; c_R\, \|X-B\|^{\beta} \tag{2.18}$$

and

$$\|(I-R)(X)\| \;\leq\; c_R\, \|X-B\|^{\beta} + \varepsilon_B, \tag{2.19}$$

provided that $\alpha\beta > 1$. Then, the order of convergence of the truncated iterative process (2.8) becomes $\alpha\beta$. However, all truncation operators used in this paper satisfy the conditions with $\beta = 1$.

Note that condition (2.9) has a clear geometrical background. If

$$R(X) := \operatorname{argmin}\{\|X-Y\| : Y \in S\}$$

is a best approximation to X in the given norm, inequality (2.9) holds with $c_R = 1$, since $B \in S$. Therefore, (2.9) with $c_R \geq 1$ can be viewed as a *quasi-optimality condition.* If the norm is defined by a scalar product, then S is a subspace, $R(X)$ is the orthogonal projection onto S and (2.9) is obviously fulfilled with $c_R = 1$.

The requirement $\alpha > 1$ for the order of convergence implies convergence in a suitable neighbourhood of B. For linear convergence ($\alpha = 1$) the additional requirement $c_\Phi < 1$ is essential.

Remark 2.6 *In the case of $\alpha = 1$ (i.e., linear convergence), the truncated process retains linear convergence, provided that $(c_R + 1)c_\Phi < 1$.*

3 Truncation operators

Theorems 2.2 and 2.4 can be applied to various classes of structured matrices. When constructing a truncation operator for a particular class, we should take care that condition (2.9) is satisfied.

3.1 General framework

Next we describe a general framework which seems to cover all important cases.

Lemma 3.1 *Let $B = R(B)$ be fixed and assume that R is Lipschitz at B. Then the inequality (2.9) holds.*

Proof. The Lipschitz property of R means that $\|R(X) - R(B)\| \leq c\,\|X - B\|$ for some constant $c > 0$ independent of X. The estimate

$$\|X - R(X)\| \underset{B=R(B)}{=} \|(X-B) + (R(B) - R(X))\| \leq (1+c)\,\|X-B\|$$

shows (2.9) with $c_R = 1 + c$. ■

Corollary 3.2 *Condition (2.9) is fulfilled as soon as $B = R(B)$ and R is a bounded linear operator.*

Let $V = \mathbb{R}^{I\times I}$ be the space of square matrices with respect to the index set I and $S \subset V$ a subspace with a prescribed sparsity pattern $P \subset I \times I$, i.e., $X \in S$ if and only if $X_{ij} = 0$ for all $(i,j) \notin P$. A familiar example of a truncation in this case is $R(X)$ defined entry-wise by

$$R(X)_{ij} := \begin{cases} X_{ij} & \text{for } (i,j) \in P, \\ 0 & \text{for } (i,j) \notin P. \end{cases} \tag{3.1}$$

This R is linear, and hence, satisfies the hypotheses of Lemma 3.1 via Corollary 3.2.

There are only rare examples, for which A and $B = f(A)$ can simultaneously be approximated by sparse matrices from $S := \{X \in \mathbb{R}^{I\times I} : R(X) = X\}$. However, it is well-known that after a discrete wavelet transform $X \mapsto L(X) := T^{-1}XT$ one can apply a matrix compression (see [7, 8, 26, 27]). Such a matrix compression is of the form (3.1) and will be denoted by Π instead of R. Then, the trunction R applied to the original matrix X is the composition of the wavelet transform L, the pattern projection Π and the back-transformation L^{-1}:

$$R := L^{-1} \circ \Pi \circ L. \tag{3.2}$$

The same product form of R is typical as well for many other choices of L and Π.

In the following lemmata the operator Π may be nonlinear.

Lemma 3.3 *Let V and W be normed spaces and $L : V \to W$ a bounded linear operator with a bounded inverse. Given $B \in V$, assume that $\Pi : W \to W$ satisfies*

$$\|Z - \Pi(Z)\| \le c_\Pi \,\|Z - L(B)\| \qquad \textit{for all } Z \in W \textit{ with } \left\|L^{-1}(Z) - B\right\| \le \varepsilon_\Phi. \tag{3.3}$$

Then the truncation operator R of the form (3.2) satisfies condition (2.9) with

$$c_R := c_\Pi \,\|L\|\, \|L^{-1}\|. \tag{3.4}$$

Proof. Let $Z = L(X)$. Then, obviously,

$$\|R(X) - X\| = \left\|L^{-1}(\Pi(Z) - Z)\right\| \le c_\Pi \|L^{-1}\|\, \|Z - L(B)\|\,,$$

and it remains to observe that $\|Z - L(B)\| = \|L(X) - L(B)\| \le \|L\|\,\|X - B\|\,.$ ■

Applications of Lemma 3.3 (especially in the case of hierarchical block matrices) are facilitated by the following construction. Define a suitable system of normed spaces $W_1, \dots, W_N$ and set

$$W := W_1 \times \ldots \times W_N = \{H = (H_1, \ldots, H_N) : H_i \in W_i\} \quad \text{with } \|H\| = \sqrt{\sum_{i=1}^{N} \|H_i\|^2}. \tag{3.5}$$

Let each W_i be associated with a truncation operator $\Pi_i : W_i \to W_i$ satisfying

$$\|H_i - \Pi_i(H_i)\| \le c_i \,\|H_i - Z_i\| \qquad \text{for all } H_i \in W_i \text{ and } 1 \le i \le N, \tag{3.6}$$

where $Z_i \in W_i$ are some fixed elements.

Lemma 3.4 *Let W be the normed space from (3.5) and let the truncation operators Π_i satisfy (3.6), where the elements $Z_i \in W_i$ are defined by*

$$L(B) = (Z_1, \dots, Z_N).$$

The product of the truncation operators Π_i defines $\Pi : W \to W$ via

$$\Pi(H) := (\Pi_1(H_1), \dots, \Pi_N(H_N)) \qquad \text{for } H = (H_1, \dots, H_N),\ H_i \in W_i.$$

Then R from (3.2) satisfies (2.9).

Proof. Let $L(X) = H = (H_1, \dots, H_N)$. Then, according to the definitions of L and Π,

$$\|H - \Pi(H)\| \le \sqrt{\sum\nolimits_{i=1}^{N} c_i^2 \, \|H_i - Z_i\|^2} \le \left(\max_{1\le i\le N} c_i\right) \sqrt{\sum\nolimits_{i=1}^{N} \|H_i - Z_i\|^2},$$

which proves (3.3) and allows us to use Lemma 3.3. ■

An important example of Π in the case of a matrix space W is given by optimal low-rank approximations.

Lemma 3.5 *Let W be a normed space of all matrices of a fixed size and let $S \subset W$ consist of all matrices whose rank does not exceed r. Then for any $H \in W$ there exists a matrix $T \in S$ such that $\|H - T\| = \min\limits_{\operatorname{rank} Z \le r} \|H - Z\|$.*

Proof. Consider a minimising sequence $Z_k \in S$, i.e., $\lim\limits_{k\to\infty} \|H - Z_k\| = \delta := \inf\limits_{\operatorname{rank} Z \le r} \|H - Z\|$. Obviously, the sequence Z_k is bounded. Therefore, a convergent subsequence $Z_{k_i} \to T$ exists. Its limit satisfies $\|H - T\| = \delta$.

The assertion $T \in S$ is due to the fact that a matrix of rank equal to $p > r$ possesses a vicinity wherein any matrix is of rank $\ge p$. ■

The optimal approximant T is not necessarily unique. For the mathematical definition of $\Pi(H)$ we choose any of the optimal approximants. In practice, the result depends of the implementation.

Corollary 3.6 *For any norm, the truncation operator Π defined in Lemma 3.4 satisfies (3.3) with $c_\Pi = 1$.*

Proof. In the given norm, no matrix in S can be closer to H than $\Pi(H)$. ■

Matrix theory provides well-developed tools for the construction of low-rank approximations in the case of any unitarily invariant norm. For an arbitrary matrix $H \in W$, denote its singular values by $\sigma_1(H) \ge \sigma_2(H) \ge \dots$ and let $\Sigma(H) := \operatorname{diag}\{\sigma_1(H), \sigma_2(H), \dots\}$. Let $\Sigma_r(H)$ be obtained from $\Sigma(H)$ by retaining all $\sigma_k(H)$ for $1 \le k \le r$ and changing the other entries into zeroes. Let $H = Q_1 \Sigma(H) Q_2$ be the singular value decomposition of H (with unitary Q_1 and Q_2). Then

$$\Pi(H) := Q_1 \Sigma_r(H) Q_2 \tag{3.7}$$

is the best possible approximant to H in the set S of matrices of rank $\leq r$, where the norm is arbitrary but unitarily invariant. It can be readily deduced from the Mirsky theorem (cf. [3, 30]) claiming that

$$\|\Sigma(H) - \Sigma(Z)\| \leq \|H - Z\| \tag{3.8}$$

for all matrices H and Z of the same size and any unitarily invariant norm. If $Z \in S$, then, clearly, $\sigma_i(Z) = 0$ for $i \geq r+1$. Using this together with the monotonicity of unitarily invariant norms (cf. [30]), we obtain

$$\|H - \Pi(H)\| = \|\Sigma(H) - \Sigma_r(H)\| \leq \|\Sigma(H) - \Sigma(Z)\|\,,$$

and, due to the Mirsky theorem, the latter norm is estimated from above by $\|H - Z\|$.

For the most familiar unitarily invariant norms such as the spectral and the Frobenius norm, the above facts can be established through simpler arguments. In particular, it is well-known that

$$\min_{\operatorname{rank} Z \leq r} \|H - Z\|_2 = \sigma_{r+1}(H), \qquad \min_{\operatorname{rank} Z \leq r} \|H - Z\|_{\mathrm{F}} = \sqrt{\sum_{i \geq r+1} \sigma_i^2(H)}\,.$$

Thus, the truncation property (2.9) is easy to achieve when a best approximation element is existing. Sometimes (e.g., for three-way approximations of bounded tensor rank) this is not the case. Nevertheless, all cases are supported by Theorem 2.4 as we can always capitalise on a quasi-optimal construction as follows.

Let $\delta(H) = \inf\limits_{T \in S} \|H - T\|$. For a given fixed $\varepsilon > 0$, let $\Pi(H)$ denote an ε-optimal approximation to H in the sense that

$$\delta(H) \;\leq\; \|H - \Pi(H)\| \;\leq\; \delta(H) + \varepsilon\,.$$

Lemma 3.7 *If $\Pi(H)$ is defined as an ε-optimal approximation to H on S, then*

$$\|H - \Pi(H)\| \;\leq\; \|H - Z\| + \varepsilon \qquad \text{for any } Z \in S\,. \tag{3.9}$$

Proof. Use $\|H - \Pi(H)\| - \|H - Z\| \underset{\|H-Z\| \geq \delta(H)}{\leq} (\delta(H) + \varepsilon) - \delta(H) = \varepsilon$. ■

In the next sections, we discuss some details of the construction of L and Π for hierarchical block matrices and matrices in the tensor format.

Other useful applications of the same general framework are Toeplitz-like matrices, where $L(X) := PX - XQ$ for some specially chosen fixed matrices P and Q (cf. [5, 28, 26]).

3.2 Application to hierarchical block matrices

Let V be the space $\mathbb{R}^{n \times n}$ of $n \times n$ matrices. Consider a block decomposition as depicted in Figure 3.1. Let N be the number of matrix blocks. Then each matrix block belongs to a certain matrix space W_i $(1 \leq i \leq N)$. Given $X \in V$, let $L_i(X) \in W_i$ be the ith block. The space W is defined according to (3.5).

The above-considered operator $L : V \to W$ maps a matrix X into the N-tuple of matrix-blocks:

$$L(X) := (L_1(X),\, \ldots,\, L_N(X)).$$

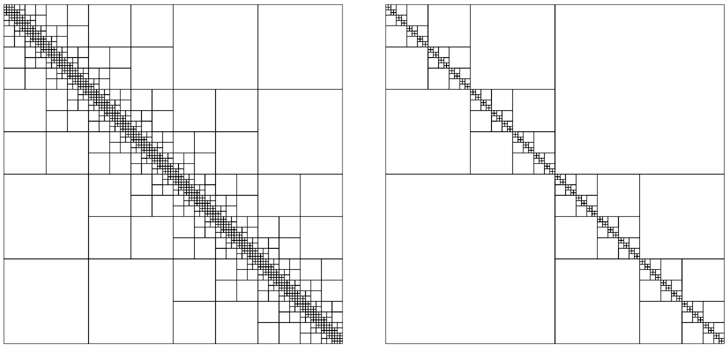

Figure 3.1: Standard and weakly admissible $\mathcal{H}$-partitionings.

If the Frobenius norm is used on the spaces V and $W_1, \ldots, W_N$, the norm induced on W is again the Frobenius norm. Obviously, $\|X\|_{\mathrm{F}} = \|L(X)\|_{\mathrm{F}}$ holds. Hence, the inverse L^{-1} exists and satisfies

$$\|L\| = \|L^{-1}\| = 1.$$

Fix a positive integer r and let $S_i \subset W_i$ be the subset of matrices of rank $\leq r$. Define S as the Cartesian product

$$S = S_1 \times \ldots \times S_N \subset W$$

and let $\Pi_i : W_i \to S_i$ be of the form (3.7) involving the singular value decomposition of the matrix block W_i. Defining $\Pi : W \to S$ as in Lemma 3.4 and using Lemma 3.5, we can apply Theorem 2.2 to $R = L^{-1} \circ \Pi \circ L$.

Note that exactly this kind of truncation is used in the theory of hierarchical block matrices (cf. [18, 19, 33, 34]) and even in some early implementations (cf. [13]).

Initially, the main purpose of the rank truncation was the reduction of storage and of the matrix-by-vector complexity. In the sequel, it was shown that with an appropriate block decomposition the hierarchical matrix structure supports all matrix operations and therefore allows to compute various matrix functions $f(A)$ of $A \in S \subset V$, where $B := f(A)$ is known to be close to S (e.g., for $f(A) = A^{-1}$ compare [1, 12], and for $f(A) = \mathrm{sign}(A)$ see [11, 22]). In spite of the observation that these computations are efficient and robust, the rigorous analysis of the intermediate truncation errors was incomplete. Our results now suggest some general framework for such an analysis of basic iterative algorithms.

Finally we remark that sometimes the optimal truncation is replaced by an approximate or heuristic one which is cheaper to compute (e.g., by cross approximation techniques, see [14, 35]). However, the rigorous analysis of such kind of quasi-optimal truncation procedures is beyond the scope of our paper.

3.3 Application to tensor approximations

Let $V_1 = \mathbb{R}^{p\times q}$ and $V_2 = \mathbb{R}^{r\times s}$, while $V = \mathbb{R}^{pr\times qs}$ for some integers p, q, r, s. The Kronecker product is a mapping from $V_1 \times V_2$ into V. For $A \in V_1$ and $B \in V_2$, the Kronecker product

$A \times B$ is defined by the block matrix $\begin{bmatrix} a_{11}B & a_{21}B & \dots \\ a_{12}B & a_{22}B & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \in V$. We say that a matrix $M \in V$ has a Kronecker rank $\leq k$, if there is a representation

$$M = \sum_{\nu=1}^{\ell} A_\nu \times B_\nu \qquad \text{with } A_\nu \in V_1,\ B_\nu \in V_2, \text{ and } \ell \leq k. \tag{3.10}$$

We define the subset of structured matrices S by the set of all matrices of Kronecker rank $\leq k$. If k is not too large, this is an interesting representation since matrices of the large size $pr \times qs$ can be described by matrices A_ν, B_ν of relatively small size.

As described, e.g., in [24], there is a simple isomorphism L from $V = \mathbb{R}^{pr\times qs}$ to $\mathbb{R}^{pq\times rs}$ such that the representation (3.10) of $M \in S \subset V = \mathbb{R}^{pr\times qs}$ is equivalent to $rank(L(M)) \leq k$. Therefore, we obtain the situation of Lemma 3.5 with $W := \Psi(V) = \mathbb{R}^{pq\times rs}$. The truncation operator is again of the form $R = L^{-1} \circ \Pi \circ L$, where Π is the optimal SVD-based truncation or an appropriate substitute.

The framework of this paper can be applied also to the (multi-linear) tensor representation (3.10) where the number of factors is greater than 2. In this case the truncation procedures are not so well developed; however, some algorithms are available and claimed to be efficient in particular applications (mostly for data analysis in chemometrics, physicometrics, etc.; cf. [6, 25]).

4 Examples of approximate iterations

We will consider iterative schemes to compute the matrix-valued functions $f(A) = A^{-1}$, $f(A) = \sqrt{A}$ and $f(A) = \mathrm{sign}(A)$. The common feature of the considered iterative schemes is that they have locally quadratic convergence and require only matrix-matrix products in each step of the iteration. We prove that our general results can be applied in the case of hierarchical matrices, Kronecker products or mixed hierarchical Kronecker-product formats to compute A^{-1} (cf. §4.1) and $\sqrt{A}$ (cf. §4.2.2).

On the other hand, our convergence theory for truncated iteration does not apply, in general, to the case of Newton-type iterative schemes in a subspace. However, numerical examples (which will be presented elsewhere) demonstrate desired convergence rate (cf. the discussion in §4.3).

4.1 Newton-Schulz iteration for calculating A^{-1}

Let $V = \mathbb{C}^{n\times n}$ and $A \in V$ a regular matrix. The Newton method applied to the equation $\Psi(X) := A - X^{-1} = 0$ yields the iteration

$$X_k := X_{k-1}(2I - AX_{k-1}) \qquad (k = 1, 2, \dots), \tag{4.1}$$

which is also named Schulz iteration (cf. [29]). This corresponds to the formulation (2.1) with

$$\Phi_k(X) := \Phi(X) := X(2I - AX).$$

The Newton method is known to have locally quadratic order of convergence (i.e., $\alpha = 2$ in (2.3)). Let $E_k := I - AX_k$ denote the error. Using $X_k = X_{k-1}(I + E_{k-1})$ we obtain

$$E_k = I - AX_{k-1}(I + E_{k-1}) = I - (I - E_{k-1})(I + E_{k-1}) = E_{k-1}^2. \tag{4.2}$$

Applying (4.2) recursively, we find that

$$E_k = E_0^{2^k} \qquad (k = 1, 2, \ldots) \tag{4.3}$$

and conclude

$$A^{-1} - X_k = A^{-1}E_k = A^{-1}E_0^{2^k} = X_0(I - E_0)^{-1}E_0^{2^k}.$$

Hence, the iteration converges quadratically for all starting values X_0 with $\rho(E_0) < 1$, where ρ is the spectral radius. Finally, equation (4.2) implies

$$A^{-1} - X_k = A^{-1}E_k = (A^{-1} - X_{k-1})A(A^{-1} - X_{k-1}),$$

which proves (2.3) with $\alpha = 2$ and $c_\Phi = \|A\|$.

Now Theorem 2.4 can be applied with a proper choice of the subset S and of the truncation operator R.

4.2 Newton iteration for the calculation of $\sqrt{A}$

4.2.1 Non-constrained Newton iteration

We apply the Newton method to the equation $\Psi(X) := A - X^2 = 0$. Abbreviating the correction by $\Delta_k := X_k - X_{k-1}$, we obtain the iteration

$$X_0 \in V, \quad X_{k-1}\Delta_k + \Delta_k X_{k-1} = A - X_{k-1}^2 \qquad (k = 1, 2, \ldots), \tag{4.4}$$

corresponding to the choice $\Phi_k(X) := \Phi(X)$, where $\Phi(X)$ solves the matrix equation

$$X(\Phi(X) - X) + (\Phi(X) - X)X = A - X^2. \tag{4.5}$$

A simple calculation shows that the latter equation implies (with the substitution $A = B^2$)

$$X(\Phi(X) - B) + XB - X^2 + (\Phi(X) - B)X + BX - X^2 = B^2 - X^2,$$

which leads to the matrix Lyapunov equation with respect to $Y = \Phi(X) - B$,

$$XY + YX = (B - X)^2.$$

Making use of the solution operator for the Lyapunov equation [11] (and assuming that $X = X^\top$ is positive definite), we arrive at the norm estimate

$$\|\Phi(X) - B\| = \left\| \int_0^\infty e^{-tX}(B - X)^2 e^{-tX} dt \right\| \le C\|B - X\|^2.$$

This proves relation (2.3) with $\alpha = 2$. Hence, Theorem 2.4 applies to the truncated version of the nonlinear iteration (4.4).

4.2.2 Newton iteration in the subspace

Let A be diagonalisable, i.e., $A = T^{-1}D_A T$ for some $T \in V$ and a non-negative diagonal matrix D_A. This gives rise to the subspace

$$V_T := \{M \in \mathbb{R}^{n\times n} : M = T^{-1}DT,\ D \text{ is diagonal}\} \subset V. \tag{4.6}$$

Note that $A \in V_T$ and that all matrices from V_T commute.

We reconsider iteration (4.4) under the assumption $X_0 \in V_T$ (this is trivially satisfied for all multiples $X_0 = a_0 A$). Next, it is easy to see that all iterates X_k of (4.4) belong to V_T. In particular, $X \in V_T$ implies $\Phi(X) \in V_T$ and the left-hand side in (4.5) can be simplified to $2X\Phi(X) - 2X^2$. Hence we obtain the iteration

$$X_0 = a_0 A, \quad X_k := \frac{1}{2}(X_{k-1} + X_{k-1}^{-1}A) \qquad (k = 1, 2, \ldots)\,, \tag{4.7}$$

where $a_0 > 0$ is the given constant. This corresponds to the formulation (2.1) with

$$\Phi_k(X) := \Phi(X) := \frac{1}{2}(X + X^{-1}A).$$

Note that newly defined Φ is different from Φ in (4.5), but both coincide on V_T. In particular for starting values $X_0 \in V_T$ both (exact) iterations yield the same X_k. Hence, the convergence analysis of §4.2.1 implies the same kind of convergence for the iteration (4.7).

For general initial values X_0 outside of V_T no quadratic convergence can be proved. Furthermore, the truncations R which we have in mind do not map V_T onto $S \cap V_T$. Therefore the truncated version of the iteration (4.7) yields approximants Y_k which not necessarily belong to V_T. Consequently, Theorem 2.4 does not apply. Nevertheless, numerical experiments with the truncated version of (4.7) show good results. In particular, the relation (2.3) holds with $\alpha = 2$ and $c_\Phi = \frac{C_0}{2}\|B^{-1}\|$.

4.3 Iterative calculation of $\operatorname{sign}(A)$

Let $A \in V = \mathbb{C}^{n\times n}$ be a matrix whose spectrum $\sigma(A)$ does not intersect the imaginary axis. The matrix function $f(A) = \operatorname{sign}(A)$ is defined by

$$\operatorname{sign}(A) := \frac{1}{\pi i}\int_{\Gamma_+}(zI - A)^{-1}\mathrm{d}z - I \tag{4.8}$$

with Γ_+ being any simply connected closed curve in $\mathbb{C}$ whose interior contains all eigenvalues of A with positive real part.

A possible iterative scheme approximating $B = \operatorname{sign}(A)$ is

$$X_0 := A/\left\|A\right\|_2, \quad X_k := X_{k-1} + \frac{1}{2}\left[I - (X_{k-1})^2\right]X_{k-1} \qquad (k = 1, 2, \ldots) \tag{4.9}$$

corresponding to $\Phi(X) := X + \frac{1}{2}\left(I - X^2\right)X$ in (2.1). This scheme has already been successfully applied in many-particle calculations (cf. [2]).

To include our scheme into the framework (2.1), we assume that A is diagonalisable, i.e., $A = T^{-1}D_AT$ with a diagonal matrix $D_A = \mathrm{diag}\{d_1, ..., d_n\}$. Again, we denote the corresponding subspace of matrices by V_T from (4.6).

Note that X_0 and all subsequent X_k belong to V_T and thus commutes with B. Taking into account that $B^2 = I$, we obtain for $X \in V_T$ that

$$\begin{aligned}\Phi(X) - B &= X - B + \frac{1}{2}(B^2 - X^2)X = (X - B)(B^2 - \frac{1}{2}(B + X)X) \\ &= \frac{1}{2}(X - B)(B(B - X) + (B - X)(B + X) = -(X - B)^2(B + \frac{1}{2}X).\end{aligned}$$

This proves the relation (2.3) with $\alpha = 2$ and $c_\Phi = \frac{3}{2} + \frac{1}{2}\varepsilon_\Phi$ and but with $X \in V$ replaced by $X \in V_T$. Hence, Lemma 2.1 leads to local quadratic convergence of the exact iteration performed in the subspace V_T.

For the global error analysis assume $A = T^{-1}D_AT \in V_T$ with $D_A = \mathrm{diag}\{d_1, ..., d_n\}$, where $d_i \in \mathbb{R}\backslash\{0\}$. As mentioned above, $X_0 \in V_T$ implies $X_k \in V_T$ for all k. The diagonal entries of D_{X_0} from $X_0 = A/||A||_2 = T^{-1}D_{X_0}T$ satisfy $d_i \in [-1, 1]\backslash\{0\}$. We have to show that the scalar iteration

$$x_k = \varphi(x_{k-1}) := x_{k-1} + \frac{1}{2}\left(1 - x_{k-1}^2\right)x_{k-1} \qquad \text{with } x_0 \in [-1, 1]\backslash\{0\}$$

converges quadratically to $\mathrm{sign}(x_0)$. Note that $\varphi(x) = xg(x)$ with $g(x) := 1 + \frac{1}{2}(1 - x^2)$. The function $\varphi : [-1, 1] \to \mathbb{R}$ is increasing and has the fixed points $\{-1, 0, 1\}$. Since $g(x) > 1$ for $x \in (-1, 1)$, we have

$$\begin{array}{ll} 0 < x_{k-1} < x_k \leq 1 & \text{if } x_0 \in (0, 1], \\ -1 \leq x_k < x_{k-1} < 0 & \text{if } x_0 \in [-1, 0). \end{array}$$

Hence, both $x = -1$ and $x = 1$ are stable fixed points.

Let $x_0 > 0$ be an initial value with $|x_0| < 1/2$. For $x \in [-1/2, 1/2]$ we have $g(x) \geq q := 11/8 > 1$, thus the number of iterations $x_k = x_{k-1}g(x_{k-1})$ to achieve a value $x_k > 1/2$ is $\mathcal{O}(\log x_0)$. Assume that $||A||_2/\rho(A) \leq \mathcal{O}(1)$. Then the smallest eigenvalue of $S_0 = A/||A||_2$ is $\mathcal{O}(1/\mathrm{cond}_2(A))$. Since this is the worst case for x_0, we obtain $\mathcal{O}(\log(x_0)) \leq O(\log \mathrm{cond}_2(A^{-1}))$ for the total number of iterations.

For $x_k \geq 1/2$, quadratic convergence is visible. In fact, $1 - x_k = \frac{1}{2}(1 - x_{k-1})^2(x_{k-1} + 2)$ implies

$$|1 - x_k| \leq \frac{3}{2}(1 - x_{k-1})^2.$$

To achieve the accuracy $\varepsilon > 0$, one requires $\mathcal{O}(\log_2 \log_2 \varepsilon^{-1})$ iterations.

Again, usual truncations lead to results outside of V_T. Then the matrices X and B in condition (2.3) no longer commute which destroys the quadratic convergence.

We examine a special version of Theorem 2.2 with $\alpha = 1$ (cf. Remark 2.6). In the case of $\alpha = 1$, we obtain the relation (2.3) with $c_{\Phi,1} = 1 + c_\Phi \|X - B\|$ and c_Φ defined below. Use

$$\begin{aligned}\Phi(X) - B &= X - B + \frac{1}{2}(B - X)(B + X)X + \frac{1}{2}(XB - BX)X \\ &= -(X - B)^2(B + \frac{1}{2}X) + \frac{1}{2}[(X - B)BX - B(X - B)X] \\ &= -(X - B)^2(B + \frac{1}{2}X) \\ &\quad + \frac{1}{2}\left[(X - B)B(X - B) - B(X - B)^2 + (X - B) - B(X - B)B\right].\end{aligned}$$

For unitary T (cf. (4.6)) we have $\|B\| = 1$ and derive the expected estimate:

$$\|\Phi(X) - B\| \leq c_\Phi \|X - B\|^2 + \frac{1}{2}\|X - B - B(X - B)B\|.$$

Hence, also Remark 2.6 does not apply, and a refined truncation error analysis for iteration (4.9) is required. However, we note that the numerical results (which will be published elsewhere) demonstrate fast and robust convergence of truncated iterations (4.9) applied to the discrete elliptic operator.

Remark 4.1 *An alternative approach can be based on the quadrature approximation to the integral (4.8) (cf. [11]). Then each matrix resolvent can be represented by the Newton-Schulz iteration discussed in §4.1.*

Another alternative is based on the observation that for nonsingular Hermitian matrices the so-called polar decomposition holds,

$$\sqrt{A^*A} = A \operatorname{sign}(A). \tag{4.10}$$

Hence, the truncated iterations from Sections 4.1, 4.2 can be directly adapted to the representation

$$\operatorname{sign}(A) = \sqrt{A^*A}\, A^{-1}. \tag{4.11}$$

Another iteration for computing sign(A) is

$$X_0 := A, \quad X_{k+1} := \frac{1}{2}(X_k + X_k^{-1}) \qquad (k = 1, 2, \ldots). \tag{4.12}$$

It converges locally quadratically to sign(A). This method is proved to be efficient in the hierarchical matrix arithmetics (see [17]).

5 Concluding remarks

In this paper we proposed a unified framework for the analysis of *truncated iterations*. The advantage of these truncated iterations is that they preserve the data-sparse structure of the intermediate matrices. The main result is that an iterative process for the evaluation of $f(A)$ can be transformed, under astonishingly general assumptions, into an implementable process which preserves the convergence rate and benefits from the underlying structure during the iterations. It is shown how this result applies to matrices in the tensor format with a bounded tensor rank and to the hierarchical matrices (with a bounded rank of the blocks).

References

[1] M. Bebendorf and W. Hackbusch: *Existence of $\mathcal{H}$-matrix approximants to the inverse FE-matrix of elliptic operators with L^∞-coefficients*, Numer. Math. **95** (2003), 1-28.

[2] G. Beylkin, N. Coult and M.J. Mohlenkamp: *Fast Spectral Projection Algorithms for Density-Matrix Computations*, J. of Comput. Phys., v. 152 (1999) 32-54.

[3] R. Bhatia: *Matrix analysis*, Springer-Verlag, New York, 1996.

[4] D.A. Bini, E.E. Tyrtyshnikov, and P. Yalamov (eds.): *Structured matrices: recent developments in theory and computation. Advances in computation.* Nova Science Publishers, Inc., Huntington, New York, 2001.

[5] D. A. Bini and B. Meini: *Solving block banded block Toeplitz systems with structured blocks: algorithms and applications*, in [4].

[6] L. de Lathauwer, B. de Moor, and J. Vandewalle: *A multilinear singular value decomposition.* SIAM J. Matrix Anal. Appl. **21** (2000), 1253–1278.

[7] J.M. Ford and E.E. Tyrtyshnikov: *Combining Kronecker product approximation with discrete wavelet transforms to solve dense, function-related systems*, SIAM J. Sci. Comp. **25** (2003), 961–981.

[8] J.M. Ford, I.V. Oseledets, and E.E. Tyrtyshnikov: *Matrix approximations and solvers using tensor products and non-standard wavelet transforms related to irregular grids*, Russ. J. Numer. Math. and Math. Modelling. **19**, No. 2 (2004), 185–204.

[9] I.P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij: *$\mathcal{H}$-matrix approximation for the operator exponential with applications.* Numer. Math. **92** (2002), 83-111.

[10] I.P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij: *Data-sparse approximation to operator-valued functions of elliptic operators.* Math. Comp. **73**, no. 247 (2003), 1297–1324.

[11] I.P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij: *Data-sparse approximation to a class of operator-valued functions.* Math. Comp. **74** (2005), 681-708.

[12] I.P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij: *Tensor-product approximation to elliptic and parabolic solution operators in higher dimensions.* Computing **74** (2005), 131-157.

[13] S.A. Goreinov, E.E. Tyrtyshnikov, and A.Y. Yeremin: *Matrix-free iteration solution strategies for large dense linear systems.* Numer. Linear Algebra Appl. **4** (1996), 1–22.

[14] S.A. Goreinov, E.E. Tyrtyshnikov, N.L. Zamarashkin: *A theory of pseudo-skeleton approximations*, Linear Algebra Appl. **261** (1997), 1–21.

[15] L. Grasedyck: *Existence and computation of a low Kronecker-rank approximation to the solution of a tensor system with tensor right-hand side.* Computing **72** (2004), 247–265.

[16] L. Grasedyck and W. Hackbusch: *Construction and arithmetics of $\mathcal{H}$-matrices.* Computing **70** (2003), 295-334.

[17] L. Grasedyck, W. Hackbusch, and B.N. Khoromskij: *Solution of large scale algebraic matrix Riccati equations by use of hierarchical matrices*. Computing **70** (2003), 121-165.

[18] W. Hackbusch: *A sparse matrix arithmetic based on $\mathcal{H}$-matrices. I. Introduction to $\mathcal{H}$-matrices.* Computing **62** (1999), 89–108.

[19] W. Hackbusch and B.N. Khoromskij: *A sparse $\mathcal{H}$-matrix arithmetic. II. Application to multi-dimensional problems.* Computing **64** (2000), 21–47.

[20] W. Hackbusch and B.N. Khoromskij: *A sparse $\mathcal{H}$-matrix arithmetic: General complexity estimates.* J. Comp. Appl. Math. **125** (2000), 479-501.

[21] W. Hackbusch and B.N. Khoromskij: *Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part I. Separable approximation of multivariate functions.* Preprint 29, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, 2005; Computing (to appear).

[22] W. Hackbusch and B.N. Khoromskij: *Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part II. HKT representations of certain operators.* Preprint 30, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, 2005; Computing (to appear).

[23] W. Hackbusch, B.N. Khoromskij, and R. Kriemann: *Hierarchical matrices based on a weak admissibility criterion.* Computing **73** (2004), 207-243.

[24] W. Hackbusch, B.N. Khoromskij and E.E. Tyrtyshnikov: *Hierarchical Kronecker tensor-product approximations.* J. Numer. Math. **13** (2005), 119–156.

[25] I. Ibraghimov: *Application of 3-way decomposition for matrix compression.* Numer. Linear Algebra Appl. **9** (2002), 551–565.

[26] V. Olshevsky, I. Oseledets, and E.E. Tyrtyshnikov: *Tensor properties of multilevel Toeplitz and related matrices.* Linear Algebra Appl., accepted for publication (2005).

[27] I.V. Oseledets and E.E. Tyrtyshnikov: *Approximate inversion of matrices in the process of solving a hypersingular integral equation.* Comp. Math. and Math. Phys. **45**, No. 2 (2005), 302–313 (translated from *JVM i MF* **45**, No. 2 (2005), 315–326).

[28] V.Y. Pan and Y. Rami: *Newton's iteration for the inversion of structured matrices.* in [4], pp. 79-90.

[29] G. Schulz: *Iterative Berechnung der reziproken Matrix.* ZAMM **13** (1933), 57–59.

[30] G.W. Stewart and J. Sun: *Matrix perturbation theory.* Academic Press, San Diego, 1990.

[31] E.E. Tyrtyshnikov: *Tensor approximations of matrices generated by asymptotically smooth functions.* Sbornik: Mathematics **194**, No. 5-6 (2003), 941–954 (translated from *Mat. Sb.* **194**, No. 6 (2003), 146–160).

[32] E.E. Tyrtyshnikov: *Kronecker-product approximations for some function-related matrices.* Linear Algebra Appl. **379** (2004), 423–437.

[33] E.E. Tyrtyshnikov: *Mosaic ranks and skeletons.* Lecture Notes in Computer Science 1196: Numerical Analysis and Its Applications. Proceedings of WNAA-96. Springer-Verlag, 1996, pp. 505–516.

[34] E.E. Tyrtyshnikov: *Mosaic-skeleton approximations.* Calcolo **33** (1996), 47–57.

[35] E.E. Tyrtyshnikov: *Incomplete cross approximation in the mosaic-skeleton method.* Computing **64** (2000), 367–380.

[36] C.F. Van Loan and N.P. Pitsianis: *Approximation with Kronecker products.* NATO Adv. Sci. Ser E Appl. Sci. **232**, Kluwer: Dordrecht, 1993, pp. 293–314.