Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 379 (2004) 423–437
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Kronecker-product approximations for some function-related matrices ☆

Eugene Tyrtyshnikov

Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina Street, 8, Moscow 119991, Russia

Received 1 October 2002; accepted 17 August 2003

Submitted by F. Uhlig

Abstract

A new approximation tool such as sums of Kronecker products is recently found to provide a superb compression property on a series of numerical examples of quite a general nature. The purpose of the paper is explanation of this phenomenon in the form of "existence theorems" for matrix approximations of low Kronecker rank for some classes of function-related matrices including important specimens from potential theory. This lays the grounds for development of new approximation algorithms, for example, in the cases when a matrix is associated with a shift-invariant function on the Cartesian product of nonunform grids, which is of great practical interest in the solution of integral equations on plates or screens.

AMS classification: 15A12; 65F10; 65F15

Keywords: Kronecker product; Low-rank matrices; Asymptotically smooth functions

1. Introduction

Approximation of matrices by sums of Kronecker products of smaller-size matrices has been recently discussed in several papers [2,4,13] and become a promising research topic with some challenges for matrix approximation theory (especially in the case of multidimensional matrices) and encouraging application examples. In

☆ Supported by the Russian Fund of Basic Research (grant 02-01-00590) and Science Support Foundation.

E-mail address: tee@inm.ras.ru (E. Tyrtyshnikov).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2003.08.013

the simplest case, given a matrix A of order $n = pq$ we want to approximate it by another matrix, A_r, of the following structure:

$$A_r = \sum_{k=1}^{r} U_k \times V_k \approx A, \tag{1}$$

where the U_k and V_k are of order p and q, respectively. [1] If $r \ll n$ then such approximations are said to be ones of low Kronecker rank. Assume that $p = q = \sqrt{n}$ and $r \ll n$. Since A_r is easily determined through the U_k and V_k, we can store only the entries of these matrices, that is, only $2rn$ numbers. If A_r is sufficiently close to A, it may replace A in computations. In this way we obtain a new matrix compression approach with the compression factor of order r/n (the ratio of memory to store A_r using its structure to the total memory to store the original matrix).

The main purpose of this paper is formulation and proof of "existence theorems" with estimates on the Kronecker rank r and the corresponding approximation error for reasonably wide classes of practically interesting matrices.

The "existence theorems" lay the grounds for development of new efficient approximation algorithms which are not considered here but can be found in our forthcoming works [5,11]. As the reader may still expect some comments on applications and algorithms, we expose the following.

If A_r were the sum of r matrices of the form $U_k \times V_k \times W_k$, each of order p (hence, $n = p^3$), then the compression factor would become of order $r/n^{4/3}$ endowing us with a *superlinear* compression property (as n grows while r being kept on the same level or grows much slower), which is not achieved in any other compression method such as the mulipole, panel clustering, mosaic–skeleton or $\mathscr{H}$-matrices [3,7,9,10,12,15–19] (we apologize for not mentioning other relevant papers).

We should remark, yet, that the matrix–vector multiplication complexity for A_r with dense unstructured U_k and V_k is reduced less dramatically than the storage. If A_r is of the form (1) then it can be multiplied by a vector in $\mathrm{O}(rn^{3/2})$ operations, less than $\mathrm{O}(n^2)$ but greater than it could be with the above-cited alternatives. Fortunately, in this paper we consider some important classes of matrices for which the U_k and V_k can be chosen of a very special structure [11], which reduces the matrix–vector multiplication complexity to the level of one of the above-cited methods and all the more improves the compression property.

Another alternative is sparsification of the U_k and V_k matrices, for example, by the discrete wavelet transform techniques [5]. In this case the acquisition of appropriate dense U_k and V_k matrices is the first step of the enterprize. On this step we may employ the approach developed in [6,8,20] for construction of low-rank approximations in which A never appears as a full array but only a procedure enabling us to pick up any requested entry of A is used.

Numerical results of [5] and theoretical estimates of [11] show that the Kronecker-product format (1), when it can be used, leads to algorithms quite competitive with,

[1] Recall that $U \times V = [u_{ij} V]$, where $U = [u_{ij}]$.

for instance, the multipole method. In contrast with the latter, the Kronecker-product algorithms of [5] do not include any preliminary analytical work and thus can be faster on the preparation stage and easily applicable to a wider class of kernel functions. Some analytical work presented in this paper is important for the proof of "existence theorems", but never used in the computations of [5] (cf. the proposals of [6,8,20]). A nice distinctive feature of the Kronecker-product approach is an *explicitly stored matrix object* with a room for further application of various methods and algorithms of linear algebra. Concerning applications and algorithms, anyway, we refer to [5,11] and suggest that more research in this direction is still to be done.

2. Motivation

Let us begin with some numerical examples.

Consider a grid $z_1, \dots, z_n$ of n nodes in $\Omega = [0, 1] \times [0, 1]$ and assume that $A = [a_{ij}]$ is associated with a function $f = f(z', z)$ of two points $z' = (x', y')$ and $z = (x, y)$ so that

$$a_{ij} = f(z_i, z_j), \quad i, j = 1, \dots, n.$$

Suppose that $n = pq$ and let $x_1, \dots, x_p$ and $y_1, \dots, y_q$ be the nodes of two uniform grids in x and y with the steps $1/p$ and $1/q$, respectively. Then, put the points (x_p, y_q) in the lexicographical order: $z_1 = (x_1, y_1)$, $z_2 = (x_1, y_2)$ and so on.

In the cases of particular interest, f is smooth everywhere but has a singularity at $z' = z$. Since the diagonal entries can be stored independently (with no damage for the linear compression property), we may assume, for definiteness, that $f(z, z) = 0$.

As the first example, consider

$$f(z', z) = 1/\rho, \quad \rho = \sqrt{(x' - x)^2 + (y' - y)^2}. \tag{2}$$

For a prescribed ε, choose r to be the smallest such that

$$\|A - A_r\|_F \leqslant \varepsilon \|A\|_F.$$

In Table 1, we take $p = q = 20$ ($n = 400$) and report on the dependence of r on ε and the aposteriori relative Frobenius-norm error ε_r. Similar results are characteristic for many other functions, find some in Table 2.

For the same functions, the Kronecker ranks remain approximately the same if we increase n when getting to finer grids on Ω. By the data compression effect,

Table 1
Kronecker-product approximations for $f = 1/\rho$ (uniform grids)

ε	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
r	4	7	9	11	13	15
ε_r	6×10^{-3}	3×10^{-5}	5×10^{-7}	5×10^{-9}	3×10^{-11}	7×10^{-14}

Table 2
Kronecker ranks for different f (uniform grids)

f	$1/\rho^2$	$1/\rho^3$	$1/\sqrt{\rho}$	$\log\rho$
$\varepsilon = 10^{-4}$	7	6	7	7
$\varepsilon = 10^{-12}$	14	14	14	15

these results resemble what we have in the mosaic–skeleton method [19,20]. The vehicle, however, is different. Now we do without block multilevel (mosaic) matrices and need not quite intricate hierarchical constructions of the multipole and panel clustering methods.

From the practical point of view, the above examples might look confusing as A is a doubly Toeplitz matrix, because of the uniformity of the grids, and thence can be stored in a straightforward way. Moreover, it can be multiplied by a vector in $\mathrm{O}(n\log n)$ operations by the FFT techniques (see, for example, [21,22]). Nevertheless, the smallness of r is still intriguing. As a matter of fact, it has nothing to do with the doubly Toeplitz structure. To show this, consider, for instance, the following nonuniform grids:

$$x_k = y_k = \frac{1}{2} - \frac{1}{2}\cos\left(\frac{(k-1/2)\pi}{p}\right), \quad k = 1, \ldots, p.$$

As we can see from Tables 3 and 4, there are very accurate approximations of low Kronecker rank in the case of nonuniform grids, too ($p = q = 20$, $n = 400$).

The presented results are obtained in a very conservative way based on the SVD algorithm (see Section 3). Of course, application of the SVD is not feasible for really large matrices. However, for such cases we can adapt the technique developed in [6,8,20]; we have found [5] that it performs fastly and reliably, at least for the cases as above and below, in Table 5, where we record the behavior of the Kronecker ranks, estimates for ε_r produced by that very algorithm, and time (in sec.) for construction

Table 3
Kronecker-product approximations for $f = 1/\rho$ (nonuniform grids)

ε	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
r	5	9	13	16	20	24
ε_r	7×10^{-3}	7×10^{-5}	4×10^{-7}	7×10^{-9}	5×10^{-11}	3×10^{-14}

Table 4
Kronecker ranks for different f (nonuniform grids)

f	$1/\rho^2$	$1/\rho^3$	$1/\sqrt{\rho}$	$\log\rho$
$\varepsilon = 10^{-4}$	8	6	8	8
$\varepsilon = 10^{-12}$	23	22	23	22

Table 5
Kronecker ranks versus n ($f = 1/\rho$, nonuniform grids)

n	5184	11 664	26 244	59 049	132 496	298 116
r	18	18	21	22	24	26
ε_r	4×10^{-5}	1×10^{-4}	5×10^{-5}	8×10^{-5}	9×10^{-5}	1×10^{-4}
Time (s)	0.1	0.3	1.0	2.7	7.1	17.1

Table 6
Kronecker ranks for discretizations of (3)

	n	1024	4096	16 384	65 536	262 144
$\varepsilon = 10^{-4}$	r	9	9	11	12	12
$\varepsilon = 10^{-5}$	r	11	14	14	16	17
$\varepsilon = 10^{-6}$	r	15	17	20	22	24

of the corresponding approximations measured on a Pentium-1600 notebook. In all cases $p = q = \sqrt{n}$.

The above numbers alone would propel us, at least the author, to seek some explanation. However, one should realize also that the considered examples are obviously related with integral equations of potential theory on plates or screens where it is necessary, for the sake of having better accuracy, to refine grids towards the edges because we have to approximate a function with infinite normal derivatives on the boundary (see, for example, [1,14]). For example, this is so for the following integral equation [1,21]:

$$\int_0^1 \int_0^1 \frac{u(x, y)}{((x_0 - x)^2 + (y_0 - y)^2)^{3/2}} \, \mathrm{d}x \, \mathrm{d}y = f(x_0, y_0), \tag{3}$$

where the integral is understood in the sense of Hadamard. This equation arises and is successfully used, for example, in the flow problems. [2] To see more numbers, consider the cells delivered by the following graded grids:

$$x_k = y_k = \frac{1}{2} - \frac{1}{2} \cos\left(\frac{k\pi}{p}\right), \quad k = 0, \ldots, p$$

and obtain a discrete version of (3) using piece-wise constant approximations on these cells and collocation at their mid-points. What the low-Kronecker-rank approximations do with this example is recorded in Table 6.

To finish with motivation, we remark that similar results are observed for some grids logically equivalent to but different from the Cartesian product of one-dimensional grids.

[2] This kind of equation is itself a challenge as there are substantial gaps in the underlying theory [1,14].

3. Kronecker products and low-rank matrices

As is noted in [23], the Kronecker-product approximations (1) can be reduced to approximations of low rank and vice versa. This is accomplished by a certain matrix transformation.

If A is $n \times n$, denote by $\mathscr{V}(A)$ a vector of size n^2 of all the entries of A taken column by column. In case $n = pq$, consider A as a block matrix,

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1p} \\ \cdots & \cdots & \cdots \\ A_{p1} & \cdots & A_{pp} \end{bmatrix},$$

where each block is $q \times q$. Then, denote by $\mathscr{P}(A)$ a matrix of size $p^2 \times q^2$ defined as follows:

$$\mathscr{P}(A) = [\mathscr{V}(A_{11}), \mathscr{V}(A_{21}), \ldots, \mathscr{V}(A_{pp})]^{\mathrm{T}}.$$

Let $U = [u_{ij}]$ be $p \times p$ and V be $q \times q$. Then

$$A = U \times V = [u_{ij} V], \quad 1 \leqslant i, \quad j \leqslant p,$$

is $n \times n$ with $n = pq$, and it is easy to check that

$$\mathscr{P}(U \times V) = (\mathscr{V}(U))(\mathscr{V}(V))^{\mathrm{T}}. \tag{4}$$

For example, if

$$U = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}, \quad V = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix},$$

then

$$\mathscr{P}(U \times V) = \begin{bmatrix} u_{11} \\ u_{21} \\ u_{12} \\ u_{22} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} & v_{12} & v_{22} & v_{32} & v_{13} & v_{23} & v_{33} \end{bmatrix}.$$

A direct generalization of (4) reads

$$\mathscr{P}\left(\sum_{k=1}^{r} U_k \times V_k\right) = \sum_{k=1}^{r} (\mathscr{V}(U_k))(\mathscr{V}(V_k))^{\mathrm{T}}, \tag{5}$$

which means that a sum of r Kronecker products is mapped to a sum of r one-ranl matrices, and, obviously,

$$\left\| A - \sum_{k=1}^{r} U_k \times V_k \right\|_{\mathrm{F}} = \left\| \mathscr{P}(A) - \sum_{k=1}^{r} (\mathscr{V}(U_k))(\mathscr{V}(V_k))^{\mathrm{T}} \right\|_{\mathrm{F}}. \tag{6}$$

Consequently, the best Frobenius norm approximation of the form (1) can be computed by the standard SVD method applied to $\mathscr{P}(A)$.

We are ready now to proceed to theory explaining the numbers of Section 2.

4. Functions, grids, matrices

The basic part of theory pertains to separable approximations of functions with some common properties.

All the examples above are with functions $f(z', z)$ that actually depend upon ρ, the distance between $z' = (x', y')$ and $z = (x, y)$. We consider a more general situation and assume that

$$f(z', z) = F(u, v), \quad u = x' - x, \ \ v = y' - y, \tag{7}$$

where $F(u, v)$ is such that any mixed derivative

$$D^m F = \frac{\partial^k \partial^l}{(\partial u)^k (\partial v)^l} F, \quad m = k + l,$$

satisfies the inequality

$$|D^m F| \leqslant c d^m m! \rho^{g-m}, \quad \rho = \sqrt{u^2 + v^2}, \tag{8}$$

with some constants $c, d > 0$ and g. We can naturally call F a *complete asymptotically smooth* function (cf. (8) for the definition and use of asymptotic smoothness with the derivatives only in part of variables). For definiteness, let $F(0, 0) = 0$.

Consider two one-dimensional grids

$$0 < x_1 < \cdots < x_p < 1, \quad 0 < y_1 < \cdots < y_q < 1 \tag{9}$$

and put their Cartesian product nodes $z_1, \ldots, z_n$, in the following order:

$$\begin{aligned} & z_i = (x_k, y_l), \quad i = l + (k-1)q, \\ & 1 \leqslant i \leqslant n, \quad 1 \leqslant k \leqslant p, \ \ 1 \leqslant l \leqslant q, \ \ n = pq. \end{aligned} \tag{10}$$

Now, if we take up $z' = (x', y')$, $z = (x, y)$ and set $u = x' - x$, $v = y' - y$, then, in case $z' \neq z$,

$$\begin{aligned} & (u, v) \in \Pi_h, \quad \Pi_h = [-1, 1]^2 \backslash (-h, h)^2, \\ & h = \min \left\{ \min_{\substack{1 \leqslant k, k' \leqslant p \\ k \neq k'}} |x_k - x_{k'}|, \ \min_{\substack{1 \leqslant l, l' \leqslant q \\ l \neq l'}} |y_l - y_{l'}| \right\}. \end{aligned} \tag{11}$$

Square and rounded brackets are used to distinquish between closed and open intervals.

The next lemma discloses why we are interested to approximate $F(u, v)$ by a function of the form

$$F_m(u, v) = \sum_{k=1}^{m} \phi_k(u) \psi_k(v) \tag{12}$$

called a *separable function*. (The only purpose of this approximation is separation of variables, so ϕ and ψ are not assumed to be smooth or anything. However, by way of arriving at separable approximations we impose some smoothness properties on F and eventually, by pursuit of the proof, come up with smooth ϕ and ψ.)

Lemma 4.1. *Suppose the nodes* $z_1, \dots, z_n$ *are defined by* (10) *and let*

$$A = [f(z_i, z_j)], \quad 1 \leqslant i, \quad j \leqslant n, \tag{13}$$

where f *is of the form* (7). *Assume that* $F(0, 0) = 0$ *and*

$$|F(u, v) - F_m(u, v)| \leqslant \varepsilon, \quad (u, v) \in \Pi_h. \tag{14}$$

Then A admits a Kronecker product approximation

$$A_r = \sum_{k=1}^{r} U_k \times V_k \tag{15}$$

with

$$r \leqslant 2m \tag{16}$$

and the error estimate

$$\|A - A_r\|_C \leqslant \varepsilon, \tag{17}$$

where $\|\cdot\|_C$ *is the maximal absolute value of the entries.*

Proof. Let $z_i = (x_{k'}, y_{l'})$ and $z_j = (x_k, y_l)$. Then

$$f(z_i, z_j) = F(x_{k'} - x_k, y_{l'} - y_l),$$

and it is easy to observe that

$$\mathscr{P}(A) = [b_{\alpha\beta}], \quad 1 \leqslant \alpha \leqslant p^2, \quad 1 \leqslant \beta \leqslant q^2,$$

where

$$\begin{aligned} &b_{\alpha\beta} = F(x_{k'} - x_k, y_{l'} - y_l), \\ &\alpha = k + (k' - 1)p, \quad 1 \leqslant k, \ k' \leqslant p, \\ &\beta = l + (l' - 1)q, \quad 1 \leqslant l, \ l' \leqslant q. \end{aligned}$$

Then, A_r is entirely defined by $\mathscr{P}(A_r)$. Let $\mathscr{P}(A_r) = [c_{\alpha\beta}]$ and set

$$c_{\alpha\beta} = F_m(x_{k'} - x_k, y_{l'} - y_l).$$

The only entries of $\mathscr{P}(A_r)$ that might be not close to those of A comprise a submatrix located in the rows with $k' = k$ and columns with $l' = l$. By virtue of (12), the rank of this submatrix is at most m. Therefore, we can nullify these entries by subtracting from $\mathscr{P}(A_r)$ an appropriate matrix of rank at most m. The rank of the renewed $\mathscr{P}(A_r)$ is not greater than $2m$. By the results of Section 2, the dyadic (skeleton) representation of $\mathscr{P}(A_r)$ of rank r is equivalent to the Kronecker product representation of A_r with r terms. The error estimate (17) evidently follows from (14). $\square$

Of course, if $F_m(u, v)$ approximates $F(u, v)$ uniformly in $[-1, 1] \times [-1, 1]$ (not only in Π_h) then we could have $r = m$. However, uniform separable approximations are feasible to prove only on Π_h.

We proceed with an elementary observation as follows.

Lemma 4.2. *Consider $F(u, v)$ for $(u, v) \in D$ with*

$$D = \bigcup_{1 \leqslant i \leqslant s} D_i, \qquad D_i = [\alpha_i, \beta_i] \times [\gamma_i, \delta_i].$$

Assume that F has a separable approximation of rank r_i on D_i, $1 \leqslant i \leqslant s$. Then F possesses a separable approximation of rank

$$r = r_1 + \cdots + r_s$$

on the whole of D.

Now we are ready to formulate one of the main theorems.

Theorem 4.1. *Assume that $F(u, v)$ is a complete asymptotically smooth function satisfying* (8). *Let $0 < h < 1$ and Π_h be defined by* (11). *Then there exists $0 < \gamma < 1$ such that, for any $m = 1, 2, \dots$, $F(u, v)$ has a separable approximation $F_r(u, v)$ of rank r with the following properties*:

$$r \leqslant (c_0 + c_1 \log h^{-1})m, \tag{18}$$

$$\begin{aligned} &|F(u, v) - F_r(u, v)| \leqslant c_2 \gamma^m \rho^g, \\ &\rho = (u^2 + v^2)^{1/2}, \quad (u, v) \in \Pi_h, \end{aligned} \tag{19}$$

where c_0, c_1 and c_2 are some positive constants.

Proof. Take $0 < a < 1$ and a positive integer s which we specify later. Consider a "rectangular ring"

$$\mathscr{R} = [-(s+1)a, (s+1)a]^2 \backslash (-sa, sa)^2$$

consisting of $4((s+1)^2 - s^2)$ basic boxes of size $a \times a$. Let D be one of these basic boxes with center (u_0, v_0) (see Fig. 1).

Assume that $u, v \in D$ and set $\Delta u = u - u_0$, $\Delta v = v - v_0$. Then, expand F into the Taylor series and consider

$$F_m(u, v) = \sum_{k=0}^{m-1} \frac{1}{k!} \left(\Delta u \frac{\partial}{\partial \xi} + \Delta v \frac{\partial}{\partial \eta} \right)^k F(\xi, \eta) \Bigg|_{\xi = u_0, \eta = v_0}. \tag{20}$$

Using the Lagrange estimate of the remainder term and the bounds of the form (8) for the derivatives of F, we obtain

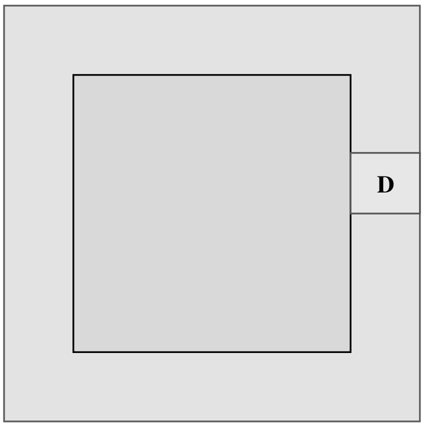

Fig. 1. A basic box D in the rectangular ring $\mathscr{R}$.

$$
\begin{aligned}
|F(u,v) - F_m(u,v)| &\leqslant (m+1)c\frac{d^m}{(sa)^m}\left(\frac{a}{2}\right)^m \mu(g) \\
&\leqslant c\left(\frac{d}{s}\right)^m \mu(g),
\end{aligned}
$$

where

$$\mu(g) = \max_{(u,v)\in D} (u^2+v^2)^{g/2}.$$

We can estimate $\mu(g)$ from above via ρ^g where $\rho = \sqrt{u^2+v^2}$ and (u,v) is an arbitrary point in D. To this end, observe that $(u,v) \in D$ implies that [3]

$$as \leqslant \rho \leqslant \sqrt{2}a(s+1).$$

In case $g \geqslant 0$,

$$\mu(g) \leqslant (\sqrt{2}(s+1)a)^g \leqslant \rho^g \left(\frac{\sqrt{2}(s+1)}{s}\right)^g$$

and in case $g < 0$,

$$\mu(g) \leqslant \frac{1}{(sa)^{|g|}} \leqslant \frac{1}{\rho^{|g|}}\left(\frac{\rho}{sa}\right)^{|g|} \leqslant \rho^g \left(\frac{\sqrt{2}(s+1)}{s}\right)^{|g|}.$$

Thus, in both cases,

$$\mu(g) \leqslant c'\rho^g, \quad c' = \left(\frac{\sqrt{2}(s+1)}{s}\right)^{|g|}. \tag{21}$$

[3] This has nothing to do with the choice of s that will be specified later.

Now, choose s so that

$$\gamma \equiv \frac{d}{s} < 1. \tag{22}$$

Then, from the above results altogether,

$$|F(u,v) - F_m(u,v)| \leqslant c_2 \gamma^m \rho^g, \quad (u,v) \in D, \quad c_2 = cc'. \tag{23}$$

To complete the proof, we cover Π_h by a family of basic boxes of different sizes, each enjoying a separable approximation of rank m and accuracy estimate of the form (23). Apparently, these can be basic boxes belonging to different rectangular rings emerging as we take

$$a = \frac{h}{s}, \left(\frac{s+1}{s}\right)\frac{h}{s}, \left(\frac{s+1}{s}\right)^2 \frac{h}{s}, \ldots.$$

The number of these rectangular rings sufficient to cover Π_h does not exceed k where

$$\left(1 + s^{-1}\right)^{k-1} \frac{h}{s} \leqslant 1.$$

Therefore,

$$k \leqslant \frac{\log h^{-1} + \log s}{\log(1 + s^{-1})} + 1.$$

Consequently, since any one of these rectangular rings consists of

$$4((s+1)^2 - s^2) = 8s + 4$$

basic boxes, the total number N of the basic boxes can be estimated as follows:

$$N \leqslant k(8s + 4). \tag{24}$$

Then, the rank estimate (16) comes up through application of Lemma 4.2. $\square$

Remark 4.1. The claim of Theorem 4.1 that the estimates (18) and (19) hold true for some $0 < \gamma < 1$ can be strengthened by observation that these estimates are valid, in fact, for any $0 < \gamma < 1$ with c_0, c_1 and c_2 depending upon γ. It follows from the proof that

$$c_0 = \mathrm{O}(\gamma^{-2} \log \gamma^{-1}), \quad c_1 = \mathrm{O}(\gamma^{-2}), \quad c_2 = \mathrm{O}(1).$$

Theorem 4.2. *Let $A = [f(z_i, z_j)]$ be a matrix of order $n = pq$, where the nodes $z_1, \ldots, z_n$ are defined by* (9) *and* (10) *and f is of the form* (7) *with a complete asymptotically smooth function F. Let*

$$h = \min \left\{ \min_{\substack{1 \leqslant k,k' \leqslant p \\ k \neq k'}} |x_k - x_{k'}|, \min_{\substack{1 \leqslant l,l' \leqslant q \\ l \neq l'}} |y_l - y_{l'}| \right\}. \tag{25}$$

Take an arbitrary $0 < \gamma < 1$. Then, for any $m = 1, 2, \ldots,$ there exists a matrix A_r of the form (15) *with r Kroneker-product terms and the following estimates*:

$$r \leqslant (c_0 + c_1 \log h^{-1})m, \tag{26}$$

$$|\{A - A_r\}_{ij}| \leqslant c_2 \gamma^m \rho_{ij}^g, \quad 1 \leqslant i, \ \ j \leqslant n, \tag{27}$$

where c_0, c_1 and c_2 are positive constants depending on γ. Here,

$$\rho_{ij} = \sqrt{(x_{k'} - x_k)^2 + (y_{l'} - y_l)^2}, \tag{28}$$

$$\begin{aligned} z_i &= (x_{k'}, y_{l'}), \quad i = l' + (k'-1)q, \ \ 1 \leqslant k' \leqslant p, \ \ 1 \leqslant l' \leqslant q, \\ z_j &= (x_k, y_l), \quad i = l + (k-1)q, \ \ 1 \leqslant k \leqslant p, \ \ 1 \leqslant l \leqslant q. \end{aligned} \tag{29}$$

Also, ρ^g is set to zero whenever $\rho = 0$.

Proof. All the premises of Theorem 4.1 are fulfilled. It remains to apply (18) and (19) and make use of Lemma 4.1. □

Remark 4.2. Estimates (26) and (27) can be reformulated as follows:

$$r = \mathrm{O}\left(\frac{\gamma^{-1}}{\log(1+\gamma)}(\log \gamma^{-1} + \log h^{-1})m\right), \tag{30}$$

$$\|A - A_r\|_C = \begin{cases} \mathrm{O}(\gamma^m h^g), & g < 0, \\ \mathrm{O}(\gamma^m), & g \geqslant 0. \end{cases} \tag{31}$$

Another view of the same claim reads: for any $\varepsilon > 0$ there exists a Kronecker-product approximation A_r of Kronecker rank

$$r = \begin{cases} \mathrm{O}(\log \varepsilon^{-1} \log h^{-1} + \log^2 h^{-1}), & g < 0, \\ \mathrm{O}(\log \varepsilon^{-1} \log h^{-1}), & g \geqslant 0, \end{cases} \tag{32}$$

and accuracy of the form

$$\|A - A_r\|_C \leqslant \varepsilon. \tag{33}$$

It is worth noting that the Kronecker rank estimates do not include n explicitly. All the same, they have something to do with n because n might be coupled with h.

In many applications, the term ρ_{ij}^g in (27) is of the magnitude of $|a_{ij}|$ and thence $\mathrm{O}(\gamma^m)$ should be regarded as the relative error bound. From the entry-wise estimates (27), it is not difficult to get to estimates in norms other than $\|\cdot\|_C$.

Note also that h is not necessary to be defined by (25). In case of an arbitrary $0 < h < 1$, it is possible to maintain (27) with a properly increased bound on r. It can be changed in the following way:

$$r \leqslant (c_0 + c_1 \log h^{-1})m + \nu,$$

where $\nu = \min(\nu_1, \nu_2)$ with ν_1 equal to the number of different values $\alpha(k', k) = x_{k'} - x_k$ such that

$$-h < \alpha(k', k) < h, \quad 1 \leqslant k', \ \ k \leqslant p$$

and ν_2 equal to the number of different values $\beta(l', l) = y_{l'} - y_l$ such that

$$-h < \beta(l', l) < h, \quad 1 \leqslant l', \ l \leqslant q.$$

It can be verified that all the above-demonstrated examples match the premises of Theorem 4.2. Thus, the exciting behaviour of the Kronecker ranks shown in Section 2 is in sheer agreement with the developed theory.

Remark 4.3. The U_k and V_k matrices appear in the above proofs, in fact, as sparse matrices, which leads to considerable reduction of the matrix–vector multiplication costs.

5. Further results

Consider now a formally different case of

$$A = [f(x_i, x_j)], \quad 0 \leqslant i, j \leqslant n, \ \ 0 < x_1 < \cdots < x_n < 1. \tag{34}$$

Let $x_i = 0.5 + (i-1)/n$ and assume that $n = pq$. Then, we may consider approximations of the form (1) as previously with r terms. Will r be small as above?

If f satisfies (7) then the answer is yes. More specifically, in this case

$$f(x', x) = \mathscr{F}(x' - x), \tag{35}$$

where, for any nonnegative integer m,

$$\left| \frac{\partial^m \mathscr{F}}{\partial x^m} \right| \leqslant c d^m m! |x|^{g-m}. \tag{36}$$

Theorem 5.1. *Let A be a matrix of the form* (34) *with f satisfying* (35) *and* (36). *Let*

$$h = \min_{1 \leqslant i, j \leqslant n} |x_i - x_j| \tag{37}$$

and assume that

$$p > h^{-1}. \tag{38}$$

Let γ be an arbitrary value such that $0 < \gamma < 1$.

Then, for any $m = 1, 2, \ldots,$ there exists a matrix A_r of the form (15) *with r Kroneker-product terms and the following estimates*:

$$r \leqslant \left(c_0 + c_1 \log \frac{1}{h - p^{-1}} \right) m, \tag{39}$$

$$|\{A - A_r\}_{ij}| \leqslant c_2 \gamma^m |x_i - x_j|^g, \quad 1 \leqslant i, \ \ j \leqslant n, \tag{40}$$

where c_0, c_1 and c_2 are positive constants depending on γ. The right-hand side of (40) *is set to zero whenever $i = j$.*

Proof. It is easy to see that

$$f(x_i, x_j) = \mathscr{F}\left(\frac{k'-k}{p} + \frac{l'-l}{pq}\right)$$

so long as

$$i = l' + (k'-1)q, \quad j = l + (k-1)q,$$

$$1 \leqslant k', k \leqslant p, \quad 1 \leqslant l', l \leqslant q.$$

Now set

$$u = \frac{k'-k}{p}, \quad v = \frac{l'-l}{q}.$$

Thus, $f(x_i, x_j)$ are the values of

$$\mathscr{F}(u + v/p) \equiv F(u, v)$$

at some grid on $-1 \leqslant u, v \leqslant 1$.

To complete the proof, we are obviously led to the need to examine separable approximations for this function $F(u, v)$. Due to (37), if

$$x_i - x_j = u + v/p \neq 0,$$

then $|u| \geqslant h - p^{-1}$. Therefore, it is sufficient to study separable approximations of $F(u, v)$ for $|u| \geqslant h_p \equiv h - p^{-1}$. To this end, we fall back to the constructions of Theorem 4.1: cover the domain $[-1, 1]\backslash(-h_p, h_p)$ by closed intervals of different size and consider the Taylor expansions for $\mathscr{F}(u_0 + \delta)$ at the central points u_0 of these intervals and $\delta = (u - u_0) + v/p$. $\square$

Remark 5.1. The same result remains for a class of nonuniform grids in x. The nodes $0 < x_1 < \cdots < x_q < 1/p$ are not necessarily equidistant but all other nodes are obtained from these by a shift of a multiple of $1/p$.

Finally we discuss possible generalizations of our results and topics for the future research.

- A direct combination of Theorems 5.1 and 4.2 can lead to useful approximations with the Kronecker products with three and more terms. Apart from sparsity and additional structure of the Kronecker-product factors, this is as well an obvious way to diminish complexity.
- Practically efficient algorithms are likely to emerge from the Kronecker-product ansatz with data-sparse (mosaic–skeleton in the spirit of papers [19,20]) approximations for the involved matrices (cf. [11]).
- We envisage that a kind of maximal-volume principle [6] can be obtained also for approximations of low Kronecker rank. Anyway, we expect that such approximations are feasible and cost-effective to compute from only a small part of entries of a given matrix (cf. [5]).

• The grids being the exact Cartesian products of smaller dimension grids is not a must. It seems plausible that similar approximation properties are retained for some grids logically equivalent to the Cartesian product grids.

References

[1] S.M. Belotserkovsky, I.K. Lifanov, Method of Discrete Vortices, CRC Press, 1992.
[2] G. Beylkin, M.M. Mohlenkamp, Numerical operator calculus in higher dimensions, PNAS 99 (16) (2002) 10246–10251.
[3] A. Brandt, Multilevel computations of integral transforms and particle interactions with oscillatory kernels, Comp. Phys. Commun. 65 (1991) 24–38.
[4] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, SIAM J. Matrix Anal. Appl. 21 (4) (2000) 1253–1278.
[5] J.M. Ford, E.E. Tyrtyshnikov, Combining Kronecker product approximation with discrete wavelet transforms to solve dense, function-related systems, UMIST, Manchester Centre for Computational Mathematics, Numerical Analysis Report No. 418, 2003.
[6] S.A. Goreinov, E.E. Tyrtyshnikov, The maximal-volume concept in approximation by low-rank matrices, Contemp. Math. 208 (2001) 47–51.
[7] S.A. Goreinov, E.E. Tyrtyshnikov, A.Y. Yeremin, Matrix-free iteration solution strategies for large dense linear systems, Numer. Linear Algebra Appl. 4 (5) (1996) 1–22.
[8] S.A. Goreinov, E.E. Tyrtyshnikov, N.L. Zamarashkin, A theory of pseudo-skeleton approximations, Linear Algebra Appl. 261 (1997) 1–21.
[9] W. Hackbusch, A sparse matrix arithmetic based on $\mathscr{H}$-matrices. I. Introduction to $\mathscr{H}$-matrices, Computing 62 (1999) 89–108.
[10] W. Hackbusch, B.N. Khoromskij, A sparse $\mathscr{H}$-matrix arithmetic. II. Application to multi-dimensional problems, Computing 64 (2000) 21–47.
[11] W. Hackbusch, B.N. Khoromskij, E.E. Tyrtyshnikov, Hierarchical Kronecker tensor–product approximations, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, Preprint No. 35, 2003.
[12] W. Hackbusch, Z.P. Nowak, On the fast matrix multiplication in the boundary elements method by panel clustering, Numer. Math. 54 (4) (1989) 463–491.
[13] I. Ibraghimov, Application of 3-way decomposition for matrix compression, Numer. Linear Algebra Appl. 9 (6–7) (2002) 551–565.
[14] I.K. Lifanov, Singular Integral Equations and Disctrete Vortices, VSP, 1996.
[15] M.V. Myagchilov, E.E. Tyrtyshnikov, A fast matrix–vector multiplier in discrete vortex method, Russian J. Numer. Anal. Math. Model. 7 (4) (1992) 325–342.
[16] V. Rokhlin, Rapid solution of integral equations of classical potential theory, J. Comput. Phys. 60 (1985) 187–207.
[17] X. Sun, N.P. Pitsianis, A matrix version of the fast multipole method, SIAM Rev. 43 (2) (2001) 289–300.
[18] E.E. Tyrtyshnikov, Mosaic ranks and skeletons, in: Proceedings of WNAA-96, Lecture Notes in Computer Science 1196: Numerical Analysis and Its Applications, Springer-Verlag, 1996, pp. 505–516.
[19] E.E. Tyrtyshnikov, Mosaic–skeleton approximations, Calcolo 33 (1–2) (1996) 47–57.
[20] E.E. Tyrtyshnikov, Incomplete cross approximation in the mosaic–skeleton method, Computing 64 (4) (2000) 367–380.
[21] E.E. Tyrtyshnikov, Toeplitz matrices, some analogs and applications (in Russian), Inst. Numer. Math., Rus. Acad. Sci. (1989).
[22] E.E. Tyrtyshnikov, A Brief Introduction to Numerical Analysis, Birkhäuser, Boston, 1997.
[23] C.F. Van Loan, N.P. Pitsianis, Approximation with Kronecker products, NATO Adv. Sci. Ser. E Appl. Sci., vol. 232, Kluwer, Dordrecht, 1993, pp. 293–314.