MOSAIC-SKELETON APPROXIMATIONS [1]

Eugene Tyrtyshnikov [2]

Abstract - If a matrix has a small rank then it can be multiplied by a vector with many savings in memory and arithmetic. As was recently shown by the author, the same applies to the matrices which might be of full classical rank but have a small *mosaic rank*. The mosaic-skeleton approximations seem to have imposing applications to the solution of large dense unstructured linear systems. In this paper, we propose a suitable modification of Brandt's definition of an asymptotically smooth function $f(x,y)$. Then we consider $n \times n$ matrices $A_n = [f(x_i^{(n)}, y_j^{(n)})]$ for quasiuniform meshes $\{x_i^{(n)}\}$ and $\{y_j^{(n)}\}$ in some bounded domain in the m-dimensional space. For such matrices, we prove that the approximate mosaic ranks grow logarithmically in n. From practical point of view, the results obtained lead immediately to $O(n \log n)$ matrix-vector multiplication algorithms.

1. **Introduction**

Mosaic-skeleton approximations arise quite naturally in many applications. Probably they were mentioned first in [10]. Later they were successfully used in the context of the boundary element method, though rather implicitly [4, 6, 7].

From the matrix analysis standpoint, a concept of *mosaic ranks* of a matrix is the one that is behind them. This concept was introduced by the author in [9].

In this paper, we describe a wide class of applications that give rise to matrices with low mosaic ranks. In particular, we propose a modification of Brandt's definition of an asymptotically smooth function $f(x,y)$. It is less restrictive than the one we considered previously [9]. In a certain sense, it seems unlikely to get any further extension. For $n \times n$ matrices $A_n =$

— Received: 2/12/1996

[1] The work was supported in part by the Russian Fund of Basic Research and also by the Volkswagen-Stiftung.

[2] Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina 8, Moscow 117333, Russia. E-mail: tee@inm.ras.ru

$[f(x_i^{(n)}, y_j^{(n)})]$, when using quasiuniform meshes $\{x_i^{(n)}\}$ and $\{y_j^{(n)}\}$ in the bounded m-dimensional domain we prove that the approximate mosaic ranks might grow only logarithmically in n.

2. There is a skeleton ...

First of all, we need to recall the definitions. A matrix of the form uv^T, where u and v are column vectors, is called a *skeleton*. If A is a sum of r linearly independent skeletons then the (classical) rank of A is equal to r. Using the skeleton expansion $A = \sum_{i=1}^{r} u_i v_i^T$ we can calculate $y = Ax$ in only $2rn$ coupled operations (instead of n^2) using only $2rn$ memory locations (instead of n^2).

In the general case, when A is $m \times n$, the compressed memory is defined by the formula

$$\text{MEMORY} = \text{RANK} \cdot (m+n).$$

For arbitrary matrices (prospectively nonsingular), the mosaic rank is introduced so that the formula

$$\text{MEMORY} = \text{MOSAIC RANK} \cdot (m+n)$$

be still valid.

If B is a submatrix for a $m \times n$ matrix A then $\Gamma(B)$ denotes the $m \times n$ matrix with the same block B and zeroes elsewhere. A system of blocks A_i is called a *covering* of A if

$$A = \sum_i \Gamma(A_i),$$

and a *mosaic partitioning* of A if the blocks have no common elements.

For any given covering, there might be far too few skeletons in every block (the fewer the better). The *mosaic rank* of A is defined as

$$\text{mr}\, A = \sum_i \text{mem}\, A_i \quad / \quad (m+n), \tag{2.1}$$

where

$$\text{mem}\, A_i = \min\, \{m_i n_i \;,\; \text{rank}\, A_i (m_i + n_i)\}. \tag{2.2}$$

The notion of the mosaic ε-rank is defined as the minimal mosaic rank over all ε-perturbations of A (usually in the 2-norm).

The use of mosaic ranks can be illuminated by the following example [9]. Let A_n be $n \times n$ with units on and below the main diagonal and zeroes in the upper triangular part. Then

$$\text{rank} A_n = n \quad \text{whereas} \quad \text{mr} A_n \le \log_2 2n.$$

Further, to cover many important applications we consider a sequence of matrices

$$A_n = [f(x_i^{(n)}, y_j^{(n)})]_{ij=1}^n,$$

of which the entries are the values of a function $f(x,y)$,

$$x \in X \subset \mathbb{R}^m, \; y \in Y \subset \mathbb{R}^m,$$

at the nodes

$$x_1^{(n)}, \ldots, x_n^{(n)} \in X, \quad y_1^{(n)}, \ldots, y_n^{(n)} \in Y.$$

Under rather general assumptions such matrices appear to be close to matrices of low mosaic rank.

Following Brandt [1], we call a function $f(x,y)$ *asymptotically smooth* if there exists g such that, for any p,

$$(2.3) \quad |\partial^p f(x,y)| \;\leq\; c_p |x-y|^{g-p},$$

where ∂^p is any p-order derivative in y (here and further on, $|x-y| = ||x-y||_2$). To be safe from infinite values, we set $f(x,y)=0$ when $x=y$.

However, we will demand a bit more of an asymptotically smooth function. In [9] we required that

$$(2.4) \quad c_p \;\leq\; c\, p^\alpha\, p! \qquad \text{for some} \qquad \alpha > 0, \; c > 0.$$

Then we considered uniform meshes

$$x_i^{(n)} = y_i^{(n)} = \frac{a}{n} i, \quad i = 1, \ldots, n.$$

and claimed that for any $\varepsilon > 0$ matrices A_n can be approximated by some $\tilde{A}_n$ such that

$$(2.5) \quad \mathrm{mr}\, \tilde{A}_n = O(\log n \log \varepsilon^{-1})$$

and

$$(2.6) \quad \|A_n - \tilde{A}_n\|_F = O(n\varepsilon).$$

That claim just shows what kind of results could be expected. The assertion itself wasn't meant to be any final, and probably it needs, all through, some correction to be valid precisely *for all* functions in question. Some applications were presented in [3]. After those preliminary results, we want now to analyze more thoroughly the premises and estimates like those of the above claim.

Here, we have some progress along the following lines: first, we relax a good deal the requirements on asymptotically smooth functions, and second, we allow the meshes to be nonuniform.

3. The basic lemma

To complete the background, we formulate the basic algebraic lemma we fall back on constantly in the derivations (for the proof, see [9]).

LEMMA 3.1. Given a matrix $A \in \mathbb{C}^{m \times n}$ of the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11} \in \mathbb{C}^{m_1 \times n_1},\ A_{22} \in \mathbb{C}^{m_2 \times n_2}.$$

assume that

$$(3.1)\quad |(m_1+n_1)-(m_2+n_2)| \le q(m+n), \quad q \ < \ 1,$$

and there exist mosaic partitionings such that

$$(3.2)\quad \operatorname{mr} A_{ii} \le c \log_2^{k+1}(m_i+n_i),\ i=1,2; \quad \operatorname{mr} A_{ij} \le r \log_2^{k}(m_i+n_j),\ i \ne j,$$

for some $k \ge 0$, and, moreover,

$$c \ge \frac{r}{\log_2 \frac{2}{q^2+1}}.$$

Then for A, there exists a mosaic partitioning such that

$$(3.3)\quad \operatorname{mr} A \le c \log_2^{k+1}(m+n).$$

4. Asymptotically smooth functions

The definition cited in Section 2 is not very satisfactory. It was okay for such an important function as $f(x,y) = 1 \ / \ |x-y|$ we worked with in [3, 9] when $x, y \in \mathbb{R}$. Still, it leaves this function out of the batch when $x, y \in \mathbb{R}^m$ for $m \ge 2$. We are thus bound to consider a more general definition.

DEFINITION 4.1. A function $f(x,y)$ is called *asymptotically smooth* if there exist $c, d > 0$ and a real number g such that

$$(4.1)\quad |\partial^p f(x,y)| \ \le \ c\, d^p\, p!\, |x-y|^{g-p},$$

where ∂^p is any p-order derivative in y.

PROPOSITION 4.1. For any m, the function

$$f(x,y) = \frac{1}{r}, \quad r = |x-y|, \quad x, y \in \mathbb{R}^m,$$

is asymptotically smooth. In particular,

$$(4.2)\quad |\partial^p \frac{1}{r}| \ \le \ \frac{4^p\, p!}{r^{1+p}}$$

for all p-order partial derivatives in y.

Since accurate estimation of derivatives is a subtle matter, to make it easier we propose an overestimation technique.

Let $\delta_i = x_i - y_i$ and take up a product

$$\delta^k = \prod_{l=1}^{k} \delta_{i_l}.$$

Any time we write δ^k it might include different indices $i_1, \ldots, i_k$. Even in this uncertainty, we have always

$$|\delta^k| \ \leq \ r^k \quad \text{and} \quad |\partial \delta^k| \ \leq \ k\, r^{k-1}.$$

If f is an algebraic sum of the form

$$f = \sum \alpha_l \, \frac{\delta^{k_l}}{r^{n_l}}$$

then, obviously,

$$|f| \ \leq \ \Gamma(f) \equiv \sum \frac{|\alpha_l|}{r^{n_l - k_l}}.$$

Now, if ϕ and ψ are two algebraic sums of (different) items of the form $\alpha \delta^k \,/\, r^n$ then let us introduce a partial ordering as follows:

$$\phi \ \prec \ \psi \quad \Longleftrightarrow \quad \Gamma(\partial^p \phi) \ \leq \ \Gamma(\partial^p \psi) \quad \forall\, \partial^p.$$

A useful observation we want to rely on consists in the following:

$$\frac{\delta^k}{r^n} \ \prec \ \frac{\delta^{k+j}}{r^{n+j}} \quad \forall\, j \geq 0.$$

Proof of the proposition. It is not difficult to see that

$$\partial^1 \frac{1}{r} \ \prec \ \frac{\delta}{r^3}$$

and

$$\partial^2 \frac{1}{r} \ \prec \ \partial \frac{1}{r^3} + \frac{1}{r^3}\, \partial \delta \ \prec \ 3\, \frac{\delta^2}{r^5} + \frac{1}{r^3} \ \prec \ 4\, \frac{\delta^2}{r^5}.$$

We now conjecture that

$$\partial^p \frac{1}{r} \ \prec \ c_p \, \frac{\delta^p}{r^{2p+1}},$$

where

$$c_p = \prod_{k=0}^{p-1} (1 + 3k),$$

and prove this by induction. It is sufficient to differentiate

$$\begin{aligned} \partial\left(\frac{\delta^p}{r^{2p+1}}\right) &\prec \frac{1}{r^{2p+1}}\,\partial\delta^p + \delta^p\,\partial\frac{1}{r^{2p+1}} \\ &\prec p\,\frac{\delta^{p-1}}{r^{2p+1}} + \delta^p\,\frac{2p+1}{r^{2p}}\,\frac{\delta}{r^3} \\ &\prec (1+3p)\,\frac{\delta^{p+1}}{r^{2(p+1)+1}}, \end{aligned}$$

and note that $c_{p+1} = (1+3p)\,c_p$. We complete the proof by passing to a rougher estimate $c_p \le p!\,4^p$. ■

Remark 4.1. As is clear from the above proof, we can provide in fact a neater estimate: for any $t > 0$,

$$c_p \;\le\; c(t)\,p!\,3^{p+t}.$$

It follows from the evident identity

$$\frac{c_p}{p!} = 3^p \prod_{k=1}^{p}\left(1 + \frac{1/3}{k}\right).$$

Using a well-known fact from the gamma-function theory:

$$\prod_{k=1}^{p}(1+\frac{x}{k}) = \frac{p^x}{\Gamma(1+x)} + O(p^{x-1}),$$

we rewrite that identity as

$$\frac{c_p}{p!} = 3^p\,\left(\frac{p^{\frac{1}{3}}}{\Gamma(\frac{4}{3})} + O(p^{-\frac{2}{3}})\right).$$

5. Why small mosaic ranks?

Denote by $C(a)$ a cube in $\mathbb{R}^m$ with the side length a, and define its far-away zone as

$$C^f(a) \equiv \{x:\ |x-y| \ge a \quad \forall\ y \in C(a)\}.$$

Take any $0 < q < 1$ and, for a while, do not change it.

Lemma 5.1. Suppose $f(x,y)$ is an asymptotically smooth function with parameters d, g. Then given any nodes such that

$$y_1, \ldots, y_l \in C(a), \qquad x_1, \ldots, x_k \in C^f(a),$$

for any positive integer p the matrix $M = [f(x_i, y_j)]_{k \times l}$ can be split as

$$M = T + R,$$

where

$$\operatorname{rank} T \;\leq\; c(d, m, q)\, p^m,$$

$$||R||_F \;\leq\; \sqrt{kl}\, q^p\, a^g.$$

PROOF. Using the Taylor expansion of $f(x, y)$ in y at some point $y_0 \in C(a)$, we may write

$$f(x, y) = T_p + R_p,$$

where

$$T_p = \sum_{k=0}^{p-1} \frac{\left((y - y_0)^T \nabla\right)^k}{k!}\, f(x, y)|_{y=y_0}.$$

It is easy to see that T_p is a sum of $O(p^m)$ *functional skeletons*, i.e., functions of the form $\phi(x)\psi(y)$. It follows that the rank of any matrix of the form $[T_p(x_i, y_j)]$ is $O(p^m)$.

For the Taylor remainder term, we obtain

$$|R_p| \leq m^p\, c\, d^p\, \frac{|y - y_0|^p}{a^p}\, a^g,$$

where m^p is an upper bound for the number of summands in the expansion of $(a_1 + \ldots + a_m)^p$, and c, d, g are taken from the definition of asymptotical smoothness. Now let us subdivide the cube $C(a)$ into s^m equal subcubes $C_1, \ldots, C_{s^m}$ with the side length $\frac{a}{s}$. If y belongs to any such cube and y_0 is its center, then for some $d_1 \geq md$

$$|R_p| \;\leq\; d_1^p\, \frac{\left(\frac{\sqrt{m}a}{2s}\right)^p}{a^p}\, a^g \leq \left(\frac{d_1\, \sqrt{m}}{2s}\right)^p\, a^g.$$

Picking up s to be the smallest possible positive integer providing us with

$$\frac{d_1\sqrt{m}}{2s} \leq q,$$

we arrive at $|R_p| \leq q^p a^g$.

We now reorder the nodes y_j to make of them a sequence of s^m clusters (maybe less, for some might be empty) coupled with every subcube, and let

$$M = [M_1, \ldots, M_{s^m}], \quad M_k = [T_p(x_i, y_j)],\; y_j \in C_k,$$

$$R = [R_1, \dots, R_{s^m}], \quad R_k = [R_p(x_i, y_j)],\ y_j \in C_k,$$

The estimates on the rank of M and Frobenius norm of R follow immediately. ■

DEFINITION 5.1. Consider a mesh $\{x_i^{(n)}\}_{i=1}^n$ and let $\mu(n, C(a))$ count the number of the nodes belonging to $C(a)$. We call a sequence of meshes *quasiuniform* on $C(a)$ if there are positive constants τ_1, τ_2 such that

$$\tau_1 \frac{b^m}{a^m}\, n \ \le\ \mu(n, C(b)) \ \le\ \max\{\tau_2 \frac{b^m}{a^m}\, n,\ 1\} \quad \forall\, C(b) \subset C(a),\ \forall\, n.$$

Throughout below we consider only those cubes of which the edges are parallel to the coordinate axes.

LEMMA 5.2. Assume that $C_1(a)$ and $C_2(a)$ are any cubes in $\mathbb{R}^m$ with no common interior points, and let $A_n = [f(x_i^{(n)}, y_j^{(n)})]_{n\times n}$, where f is an asymptotically smooth function with parameters d, g, and the sequences of meshes $\{x_i^{(n)}\}$ and $\{y_j^{(n)}\}$ are quasiuniform on $C_1(a)$ and $C_2(a)$, respectively. Then for any positive integer p and for all n there exist splittings

$$A_n = T_n + R_n$$

with

$$\operatorname{mr} T_n = O(p^m), \quad ||R_n||_F = O(n^{\gamma}\, q^p\, a^g)$$

for some $\gamma > 0$. In case $m + 1 + 2g \ > \ 0$ the latter holds with $\gamma = 1$.

PROOF. Subdivide every cube into s^m subcubes with the side length $\frac{a}{s}$. We'd like to proceed by induction, and here the choice of s is where we should be very careful (for instance, if $s = 2$ then it makes no good for the proof).

Consider the most close faces of $C_1(a)$ and $C_2(a)$ and the subcubes adjoint to them. As is readily seen, the number of pairs of which one subcube is not in the far-away zone of the other one is upper bounded by $3^m\, s^{m-1}$.

We now reorder the nodes to have sequences of clusters coupled with the subcubes. Once this is done, we recognize in A_n a block structure with $s^m \times s^m$ blocks. Due to Lemma 5.1, each block save for at most $3^m\, s^{m-1}$ among them can be split into $T + R$ where $\operatorname{rank} T = O(p^m)$ while $||R||_F = O(\frac{n}{s^m}\, q^p\, (\frac{a}{s})^g)$.

We now reduce the construction of a proper mosaic partitioning to the same problem only for the most close subcubes. The number of corresponding blocks does not exceed $3^m s^{m-1}$. Hence, we have the recursive relationship for the memory as follows:

$$\mathrm{mem}\,(n) \le O(np^m) + 3^m s^{m-1}\, \mathrm{mem}\,(\frac{n}{s^m}).$$

To make it work we need to choose s so that $\frac{3^m}{s} < 1$.

That is still only half the matter. Another half goes with the estimation of norms. When applying Lemma 5.1 recursively, we have to replace a by $\frac{a}{s^{k-1}}$ at the kth step, which finally leads to the estimate

$$||R_n||_F^2 \;\le\; c'\,(nq^p a^g)^2 \sum_{k=0}^{O(\log n)} \left(\frac{3^m\; s^{m-1}}{s^{2m}}\; s^{-2g}\right)^k, \quad c' > 0.$$

If $m + 1 + 2g \;>\; 0$, then we choose s to be the smallest possible one subject to

$$\left(\frac{3^m\; s^{m-1}}{s^{2m}}\; s^{-2g}\right) \;<\; 1.$$

■

We are now in a position to set off the main theorems.

Theorem 5.1. Given any asymptotically smooth function $f(x,y)$ and quasiuniform sequences of meshes $\{x_i^{(n)}\}$ and $\{y_j^{(n)}\}$ on a cube in $\mathbb{R}^m$, for any $\varepsilon > 0$ and for all n the matrices $A_n = [f(x_i^{(n)}, y_j^{(n)})]_{n\times n}$ can be split as

$$A_n = T_n + R_n,$$

where

$$\mathrm{mr}\, T_n = O(\log n\;\log^m \varepsilon^{-1}),$$

$$||R_n||_F = O(n^\gamma\; \varepsilon)$$

for some $\gamma > 0$. In case $m + 2g \;>\; 0$ the latter holds with $\gamma = 1$.

Proof. Given a cube $C(a)$, we subdivide it into s^m equal subcubes with the side length $\frac{a}{s}$ and then permute the nodes to make out a sequence of clusters attached to every subcube. The matrix A_n turns to be a block matrix with $s^m \times\; s^m$ blocks. With $s = 2$ we can easily come to the estimate on mosaic ranks by induction, that's where Lemma 3.1 features in. Here, we also use Lemma 5.2 for $p = \log \varepsilon^{-1}$ to be chosen.

Inductive estimation of norms might require to increase s. We reduce the problem to the same one but only for those blocks that correspond to most close subcubes. Altogether, there are no more than $3^m\ s^m$ blocks like those. Consequently,

$$||R_n||_F^2 \ \leq \ c'\ (n^\gamma q^p a^g)^2 \sum_{k=0}^{O(\log n)} \left(\frac{3^m\ s^m}{s^{2m\gamma}}\, s^{-2g} \right)^k, \quad c' > 0.$$

In case $m + 2g \ > \ 0$, this entails $m + 1 + 2g \ > \ 0$, and hence, from Lemma 5.2, $\gamma = 1$. In this case, for some s the denominator of the progression gets less than 1. Still wishing to apply Lemma 3.1, we may take s equal to a power of 2. ■

Theorem 5.2. Under the provisions of Theorem 5.1, for any $\delta > 0$ there are splittings such that

$$\mathrm{mr}\, T_n = O(\log^{m+1} n), \qquad ||R_n||_F = O(\frac{1}{n^\delta}).$$

To prove this, we've recourse to Theorem 5.1 and then choose

$$\varepsilon = \frac{1}{n^{\delta+\gamma}}.$$

I'd like to thank Nickolai Zamarashkin for useful remarks.

REFERENCES

[1] A. Brandt, *Multilevel computations of integral transforms and particle interactions with oscillatory kernels,* Computer Physics Communications **65**, (1991) 24–38.

[2] S. A. Goreinov, E. E. Tyrtyshnikov, N. L. Zamarashkin, *Pseudo-skeleton approximations of matrices,* Dokladi Rossiiskoi Akademii Nauk **343 (2)**, (1995) 151–152.

[3] S. A. Goreinov, E. E. Tyrtyshnikov, A. Y. Yeremin, *Matrix-free iteration solution strategies for large dense linear systems,* Numer. Linear Algebra Appl., (1996).

[4] 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.

[5] N. Mikhailovski, *Mosaic approximations of discrete analogues of the Calderon-Zigmund operators, Manuscript, INM RAS (1996).*

[6] M. V. Myagchilov, E. E. Tyrtyshnikov, *A fast matrix-vector multiplier in discrete vortex method,* Russian Journal of Numerical Analysis and Mathematical Modelling **7 (4)**, (1992) 325–342.

[7] V. Rokhlin, *Rapid solution of integral equations of classical potential theory,* J. Comput. Physics **60**, (1985) 187–207.

[8] E. E. Tyrtyshnikov, *Matrix approximations and cost-effective matrix-vector multiplication,* Manuscript, INM RAS (1993).

[9] E. E. Tyrtyshnikov, *Mosaic ranks and skeletons,* Lecture Notes Math., (1996).

[10] V. V. Voevodin, *On a method of reducing the matrix order while solving integral equations,* Numerical Analysis on FORTRAN, Moscow University Press, (1979) 21–26.