Rectangular maximum volume and projective volume search algorithms.

Osinsky A.I.
February 19, 2019

Abstract

New methods for finding submatrices of locally maximal volume and large projective volume are proposed and studied. Detailed analysis is also carried out for existing methods. The effectiveness of the new methods is shown in the construction of cross approximations, and estimates are also proved in the case of their application for the search for a strongly nondegenerate submatrix. Much attention is also paid to the choice of the starting submatrix.

Keywords: Low rank approximations; Pseudoskeleton approximations; Maximum volume principle

1 Introduction

In this paper, we study algorithms aimed at finding submatrices of large volume and projective volume. The volume and the projective ($r$-projective) volume for an arbitrary matrix $A \in \mathbb{C}^{M \times N}$ are defined as follows:

$$V(A) = \text{vol}(A) = \nu_2(A) = \prod_{i=1}^{\min(M,N)} \sigma_i(A),$$

$$\nu_r(A) = \prod_{i=1}^{r} \sigma_i(A).$$

The submatrices with close to the maximum volume (or projective volume) are known to yield precise cross approximations [1, 4, 12]. Especially we pay attention to the projective volume, which guarantees better accuracy estimates. It is also confirmed in the numerical experiments.

An important feature of the cross approximations is the use of a small part of matrix elements. This requires efficient algorithms for large volume (or large projective volume) submatrix search.

Since the maximum volume submatrix search is an NP-complete problem, in practice much simpler computational algorithms are used. One of such algorithms is called maxvol
It finds a square submatrix of locally maximal volume (whose volume does not increase by replacing a single row or column). Such submatrices are called "dominant". In practice, those submatrices are often in "good" rows and columns for the cross approximation. Moreover, there are estimates on the norm of the pseudoinverse matrix guaranteeing that "dominant" submatrices are relatively well conditioned.

In this paper, we consider a generalization of the maxvol algorithm, which searches for large projective volume submatrices. The main result is the construction of a search algorithm for dominant rectangular submatrices and its complete analysis. Also, a good choice of the starting submatrix is studied, which provides a restriction on the number of row or column exchanges in the algorithm.

The main idea of the proposed algorithm for large projective volume is to search for "good" rows and columns separately. Since the sought-for \( m \times n \) submatrix is expected to be "good" only in the sense of its largest \( r \) singular numbers, it is possible to start with \( r \times n \) and \( m \times r \) submatrices of large volume. They are in the "good" columns and rows, respectively. In this case, the dominance along the smaller size ensures that \( r \) rows and columns are now sufficiently "good" to further select \( m \) rows and \( n \) columns.

In the current section, we introduce the necessary notation and give lemmas concerning the volume, projective volume, and dominant submatrices.

In the 2 section the algorithms are constructed and briefly described. Detailed versions can be found in the appendix.

In the section 3 the numerical experiments on random matrices are carried out. The results confirm the hypothesis of high accuracy in the Frobenius norm.

1.1 Notations

Before proceeding further, we formulate the basic notation concerning column and cross approximations.

The original matrix is usually denoted by \( A \). Its size is \( M \times N \), unless otherwise specified. \( \hat{A} \) is used to denote a submatrix of \( A \), whose rows and columns generate CGR-approximation. The symbol \( \hat{\cdot} \) everywhere below denotes the submatrix of the corresponding matrix.

Matrices \( U \) and \( V \) everywhere further contain orthonormal rows and/or columns. In spite of the fact that they are possibly rectangular, we shall also call them "unitary". They often arise from the singular value decomposition for a matrix \( A \in \mathbb{C}^{M \times N} \) of rank \( R \) as follows:

\[
A = U \Sigma V, \\
U \in \mathbb{C}^{M \times R}, \\
\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_R), \\
\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_R, \\
V \in \mathbb{C}^{R \times N}.
\]

Then \( \hat{U} \in \mathbb{C}^{n \times R} \) and \( \hat{V} \in \mathbb{C}^{R \times n} \) denote some submatrices corresponding to \( n \) rows \( R \) and columns \( C \) of \( A \).

Everywhere below the approximation rank is denoted by \( r \).
For an arbitrary matrix $B$ of rank at least $r$, the matrices $B_r$ and $B_r^+$ (called $r$-pseudoinverse) are defined in terms of the singular value decomposition $B = U\Sigma V$ as follows:

$$B_r = V^*\Sigma_r U^*, \quad \sigma_i(\Sigma_r) = \begin{cases} \sigma_i(\Sigma), & \text{if } i \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

$$B_r^+ = V^*\Sigma_r^+ U^*, \quad \sigma_i(\Sigma_r^+) = \begin{cases} \sigma_i^{-1}(\Sigma), & \text{if } i \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

Thus $B_r$ is the best rank-$r$ approximation of $B$ with respect to the 2-norm and the Frobenius norm. If $r$ coincides with the rank of $B$, we can write $B_r^+$ (exact pseudoinverse) instead of $B_r^+$. $r$-pseudoinversion is further mainly used for submatrices of a large projective volume to construct a pseudo-skeleton approximation.

### 1.2 Properties of extreme submatrices

This subsection describes the main properties of submatrices with a locally maximal volume or projective volume, proved in [12]. Naturally, properties of the projective volume are also true for volume.

**Lemma 1 ([12])**. For submatrices $A \in \mathbb{C}^{m \times n}$ and $\hat{U} \in \mathbb{C}^{m \times n}$ with the locally maximal $r$-projective volume, the following properties hold:

1. For a matrix $W = \begin{bmatrix} A & b \end{bmatrix} \in \mathbb{C}^{m \times (n+1)}$, $r \leq n$, the inequality for the $r$-projective volume holds

   $$V_2^r(A)\sqrt{1 + \|A_r^+b\|_2^2} \leq V_2^r(W) \leq V_2^r(A)\sqrt{\frac{n+1}{n-r+1}}.$$

   This implies

   $$\|A_r^+b\|_2^2 \leq \frac{r}{n-r+1}.$$

2. Let $U \in \mathbb{C}^{m \times N}$ be a matrix with orthonormal rows: $UU^* = I_m$, and $\hat{U}$ be a submatrix of $U$. Then

   $$\|\hat{U}_r^+\|_2 \leq \sqrt{1 + \frac{r(N-n)}{n-r+1}},$$

   $$\|\hat{U}_r^+\|_r \leq \sqrt{r + \frac{r(N-n)}{n-r+1}}.$$

The second property is especially important since the cross approximation accuracy is affected by the properties of some unitary matrices' submatrices.

The next statement shows that the locally maximal volume of long rectangular submatrices is close to the globally maximal volume, which may be false for square submatrices.

**Consequence 1.** Let $\hat{A} \in \mathbb{C}^{n \times r}$ be a submatrix of $A \in \mathbb{C}^{N \times r}$ with the maximal (nonzero) 2-volume among all the submatrices, which differ from $\hat{A}$ in a single row. And $A_M \in \mathbb{C}^{n \times r}$ be a submatrix with the maximal 2-volume in $A$. Then

$$V_2(A_M) \leq V_2(\hat{A}) \cdot \left(\frac{n+1}{n-r+1}\right)^{r/2}. \quad (1)$$
Proof. In justifying the Dominant-C algorithm in section 2, we show the following. If the j-th row is replaced by i-th one \((i > n)\) then the ratio of the new volume to the old one equals
\[
\frac{V_{\text{new}}}{V_{\text{old}}} = |C_{ij}|^2 + (1 + \|C_i\|^2_2)(1 - \|C_j\|^2_2),
\]
(2)
where \(C = A\hat{A}^+\).

Since the ratio of r-projective volumes of submatrices does not change after multiplication by any matrix, now we can estimate the maximal r-projective volume in \(C\).

First, we estimate the squared 2-norm of the j-th row of \(C\), denoted by \(l_j\). Let \(x = \max_{i > n} l_i\).

Then, since the ratio in (2) does not exceed 1,
\[
\forall j = 1, n : \ (1 - l_j)(1 + x) \leq 1
\]
and therefore
\[
\forall j = 1, n : \ l_j \geq \frac{x}{1 + x}.
\]
(3)

Let \(A_M\) contain \(k > 0\) rows from \(\hat{A}\). The sum of the squared lengths of all rows in \(C\) corresponding to \(\hat{A}\) (that is, rows of \(\hat{C}\)) equals \(r\). So the squared Frobenius norm (compared to all \(l_i = x\)) increases by no more than
\[
r - (n - k) \frac{x}{1 + x} - kx = r - \frac{nx}{1 + x} - kx,\]
ssince we replace \(k\) rows with lengths \(x\) (thus \(-kx\)) with \(k\) rows from \(\hat{A}\). This accounts for the squared Frobenius norm at most
\[
r - (n - k) \frac{x}{1 + x}.
\]

Indeed, as already mentioned, the squared Frobenius norm of the entire matrix \(\hat{A}\) is equal to \(r\). On the other hand, the sum of squared 2-norms of the rest \(n - k\) rows, according to (3) is at least \((n - k)\frac{x}{1 + x}\).

The maximum of this expression is achieved at \(k = 1\). Then
\[
\|C_M\|^2_F \leq nx + r - \frac{x(n + x)}{1 + x} = r + \frac{x^2(n - 1)}{1 + x}.
\]

This expression grows with increasing \(x\). Choosing the maximum according to point 1 of the lemma \([1]\) \(x = \frac{r}{n - r + 1}\), we get
\[
\|C_M\|^2_F \leq r + \frac{r^2(n - 1)}{(n - r + 1)(n + 1)} \leq \frac{n + 1}{n - r + 1}.
\]

In the case \(k = 0\), using point 1 of the lemma \([1]\) again for \(n\) rows of \(C\), we obtain
\[
\|C_M\|^2_F \leq \frac{r}{n - r + 1},
\]
which is lower than the previous estimate.

The maximum projective volume is achieved in case of equal \(r\) singular numbers. This means they are all equal to \(\sqrt{\frac{n + 1}{n - r + 1}}\), which proves (1). \(\square\)
Remark 1. For \( n \approx r^2 \), the volume ratio is constant.

At last, we give one more significant property.

Lemma 2 ([9]). For an arbitrary matrix \( A \in \mathbb{C}^{M \times N} \) and any \( r \leq \min(M, N) \), the following identity holds

\[
\sum_{T, |T|=r} \mathcal{V}^2(A_T) = \sum_{1 \leq t_1 < t_2 < \ldots < t_r \leq \min(M, N)} \sigma_{t_1}^2 \sigma_{t_2}^2 \cdots \sigma_{t_r}^2 \geq (\mathcal{V}_2^r(A))^2,
\]

(4)

where the sum is taken over all sets of \( r \) indices \( T = \{t_1, \ldots, t_r\} \). \( A_T \) are the rows of the matrix \( A \) with indices from \( T \), and \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(M,N)} \) are the singular numbers of the matrix \( A \).

The inequality in (4) is obtained by taking into account only the first summand.

2 Algorithms

The primary intention of this section is to obtain and to justify new algorithms. Especially, an algorithm for finding dominant rectangular submatrices. For every algorithm, we estimate the iteration numbers and the computational complexity. The complete descriptions with conversion formulas are given in the Appendix.

All the basic algorithms somehow use fast recalculations to replace one of the rows or columns of the current submatrix. The replacements always increase the volume of the current submatrix. The criteria for such substitutions are given in the current section.

On the other hand, the projective volume cannot be recalculated using rank-2 updates. Therefore, a more complex approach is used.

We search for \( m \times n \) submatrix of large \( r \)-projective volume by using \( r \times n \) and \( m \times r \) submatrices of large 2-volume. The construction idea is illustrated in figure 1.

The informal justification is as follows. Suppose that \( \text{rank} \ A = r \) and some \( r \times n \) and \( m \times r \) submatrices are submatrices of maximal volume. Then it is easy to see that a submatrix \( \hat{A} \) at the intersection of the \( m \) rows and \( n \) columns is indeed a submatrix of the maximal \( r \)-projective volume.

This algorithm is called \text{maxvol-proj}. The algorithm for searching for large 2-volume submatrices is called \text{maxvol-rect} since in the general case the submatrices are rectangular.

Algorithm \text{maxvol-proj}(r, m, n)

Input: Matrix \( A \in \mathbb{R}^{M \times N} \), starting sets of row indices \( I \) and column indices \( J \) columns \( r \) and final sizes of \( C \) and \( R \) (\( n \) and \( m \), respectively).

Output: The set \( I \) of the row indices of \( R \) and the set \( J \) of the column indices of \( C \) that contain a submatrix of a large projective volume.

1. \text{maxvol-rect}(A, I, -)
2. \text{maxvol-rect}(A^T, J, -)
Thus, we reduce the problem of finding a submatrix of a large projective volume to the search for two smaller submatrices of large 2-volume.

The algorithm maxvol-rect consists of two parts. First part is a search for large 2-volume submatrix in the fixed rows. The second part is the search for large 2-volume submatrix in the fixed columns. Without loss of generality, we assume that there are more rows than columns. Otherwise, the algorithm differs only in transposition.

The rest of the section is mainly devoted to the two following algorithms. Algorithm Dominant-C operates in columns and replaces rows of a \( m \times r \) submatrix. Algorithm Dominant-R operates in rows and replaces columns. Let us first consider the existing algorithms for searching for large volume submatrices.

We start with the maxvol algorithm. It looks for the \( r \times r \) submatrix of large volume. Then we look at the maxvol2 algorithm, which greedily adds rows to maximize the 2-volume of the current submatrix.

The idea of maxvol is to find a dominant submatrix. A submatrix is called dominant if replacing any single row (column) by another row (column) of the matrix does not increase its volume. The criterion for the submatrix dominance comes from the following lemma.

**Lemma 3** (see proof of lemma 1 in [5]). Let \( \hat{A} \in \mathbb{C}^{r \times r} \) be a submatrix in the first \( r \) rows of the matrix \( A \in \mathbb{C}^{M \times r} \). Then, replacing the \( j \)-th row of the submatrix \( \hat{A} \) with the \( i \)-th row of \( A \) \( (i > r) \) changes the squared volume of \( \hat{A} \) as

\[
\frac{V_2^2}{V_1^2} = |C_{ij}|^2,
\]

where

\[
C = A\hat{A}^{-1}.
\]

It immediately follows that the \( r \times r \) submatrix in fixed \( r \) columns is dominant if and only if all absolute values of the elements in \( C \) do not exceed 1, that is, \( \|C\|_C \leq 1 \). If this inequality is not satisfied, one can replace the rows.

**Algorithm maxvol\( (r) \) [5]**

**Input:** Matrix \( A \in \mathbb{C}^{M \times N} \), starting sets of row indices \( I \) and column indices \( J \) of cardinality \( r \). For example, \( I = J = \{1, \ldots, r\} \).

**Output:** The row and column indices of the dominant submatrix of rank \( r \) written in \( I \) and \( J \).

![Figure 1: Submatrix \( \hat{A} = A_{I,J} \) returned by maxvol-proj algorithm.](image-url)
1:  C := A_{i,j} A_{i,j}^{-1}
2:  while replacements occur do
3:      for changes in \{\mathcal{I}, \mathcal{J}\} do
4:          C_{i,j} := \arg \max_\|C\|_C
5:      while |C_{i,j}| > 1 do
6:          Recalculate C
7:          Change j to i in changes_in
8:          C_{i,j} := \arg \max_\|C\|_C
9:      end while
10:  end for
11: end while

The algorithm complexity is \(O((M+N)r \cdot \text{iter} + (M+N)r^2 \cdot IT)\) operations. Here iter denotes the total number of replacements of rows and columns, that is, the number of iterations of “while” loop on the line 5 and IT is the number of passes through rows and columns, that is, the total number of iterations of the “for” loop (line 3).

In practice \(|C_{i,j}|\) is compared to \(c = 1 + \varepsilon\) for small \(\varepsilon\). This avoids instability and, as we will see below, allows to limit the number of row/column replacements at each step of the algorithm.

So, let \(|C_{i,j}|\) be compared with some constant \(c > 1\) in the line 5. In this case, the algorithm no longer produces a dominant submatrix. However, the volume of the resulting submatrix differs by no more than \(c\) times from the volume of any submatrices that differs in a single row or column.

Let us denote by \(\Gamma\) the ratio of the maximum volume to the volume of the starting submatrix. Then for \(\Gamma = (\alpha^2 r)^{r/2}\) we can guarantee the convergence within the corresponding rows or columns in

\[
k = \left\lceil \frac{\ln \Gamma}{\ln c} \right\rceil - 1
\]

steps. Indeed, the C-norm of rows of the matrix C cannot be greater than the ratio of the current volume to the maximum one, and this ratio for each iteration drops by at least \(c\) times.

If \(\alpha \leq c\), we need no more than

\[
k_1 = \left\lceil \frac{\ln c^2 r}{2 \ln c} \right\rceil - 1 \leq \left\lceil r \left(1 + \frac{\ln r}{2 \ln c}\right) \right\rceil - 1
\]
steps.

If $\alpha > c$, then the $C$-norm of some row is at least $\alpha$. This leads to the reduction in the volume ratio by $\alpha$ times. Assume $\alpha$ becomes less or equal to $c$ after $k_2$ iterations. Then we get the following condition:

$$\alpha^{(1-\frac{1}{r})^{k_2}} \leq c$$

$$k_2 = \left\lfloor \frac{\ln \frac{\ln c}{\ln \alpha}}{\ln (1-\frac{1}{r})} \right\rfloor = \left\lfloor \frac{\ln \alpha}{\ln (\frac{r}{r-1})} \right\rfloor \leq \left\lfloor r \ln \frac{\ln c}{\ln \alpha} \right\rfloor. \quad (5)$$

Adding $k_1$, in view of the rounding, we obtain

$$k \leq r \left( 1 + \frac{\ln r}{2 \ln c} + \max \left( 0, \ln \frac{\ln r^{2/r}}{\ln c} \right) \right). \quad (6)$$

Let’s consider how big the ratio $\Gamma$ can be for a randomly selected starting submatrix. Let us apply the maxvol algorithm to a matrix with randomly distributed matrices of left and right singular vectors. Since the Gaussian matrices are unitary invariant and $\Gamma$ equals the volume ratio, the current matrix $C$ can be considered Gaussian.

The probability of the volume of a random Gaussian matrix to differ from its expectation is exponentially small. Even though the maximum is taken over $M^rN^r$ different matrices, we obtain with a high probability $\Gamma = O(r^r \ln r^r MN)$. If $c = \text{const} > 1$, the substitution in (6) yields

$$k = O(r(\ln r + \ln \ln \ln MN)) \approx O(r \ln r)$$

and $\text{iter} \approx O(r \ln r \cdot \text{IT})$.

Of course, it would be better to have a guarantee on $\Gamma$. Such a guarantee can be useful in the search for a strongly nondegenerate square submatrix in a long rectangular matrix.

In order to limit the number of steps, it is sufficient to construct RRQR (rank-revealing QR) via Householder reflections. This algorithm iteratively adds a new row/column corresponding to the row/column of the current error with the largest 2-norm (for a detailed description, see [6], page 278). An alternative version of this algorithm called pre-maxvol is presented in the Appendix. This version uses only QR of the current submatrix and the matrix $C$ (as in maxvol). Thus asymptotics and result remain. Pre-maxvol is sufficient to obtain a submatrix with a volume that differs from the maximum volume submatrix by no more than $r!$ times.

Obviously, the ratio of the new volume to the old one is equal to the length of the new row in the orthogonal complement to the already selected rows. These squared lengths are further written in a vector $\gamma$.

Algorithm pre-maxvol($r$) (The full version is a modification of the addition of rows in RRQR from [7])

Input: Matrix $A \in \mathbb{R}^{M \times n}$, required rank $r$.

Output: Set of row indices $\mathcal{I}$ of cardinality $r$, containing a submatrix with volume no more than $r!$ less than the maximum.

1: $\mathcal{I} := \emptyset$
2: for \( i := 1 \) to \( M \) do
3: \( \gamma_i := \|A_i\|_2^2 \)
4: end for
5: for \( k := 1 \) to \( r \) do
6: \( \gamma_j := \arg \max \|\gamma\|_C \)
7: \( I := I \cup \{j\} \)
8: Recalculate elements of \( \gamma \): squared row lengths of \( A \) projected on the orthogonal complement to the rows in \( I \)
9: end for

The complexity of \texttt{pre-maxvol} is \( O(Mnr) \) for \( A \in \mathbb{C}^{M \times n} \). Thus for the algorithm \texttt{maxvol} in these columns (the case \( n = r \)),

\[
  k_2 \leq \left\lfloor r \ln \frac{\ln r}{\ln c} \right\rfloor.
\]

The asymptotics of the number of steps \texttt{maxvol} takes is entirely determined by \( k_1 \), so

\[
  \text{iter} = O(r \ln r \cdot \text{IT})
\]

is guaranteed when \( c = \text{const} > 1 \).

The following statement was first proved in [18], where the greedy addition of lines was also investigated. Here we give a simpler proof using the projective volume.

**Proposition 1** ([18]). The algorithm \texttt{pre-maxvol} finds a submatrix with the volume not more than \( r! \) times smaller than the maximum.

**Proof.** We yield the proof by induction on \( r \).

The base \( r = 1 \) is obvious.

To prove the induction step, we first choose the longest row, and denote it by \( c \). Then we look at the orthogonal complement to \( c \). This means a decrease of \( n \) and \( r \) by one. All the rows are further considered in the orthogonal complement to \( c \).

By the induction hypothesis, one can find a submatrix with volume that differs not more than \( (r - 1)! \) times from the maximum. But for the original matrix, a restriction holds. The maximum volume is taken among the submatrices containing the already selected row \( c \) (not the total maximum).

Now we should prove the following: among this subset, there is a submatrix with volume no more than \( r \) times less than the total maximum. Then we obtain the submatrix with the volume no more than \( r \cdot (r - 1)! = r! \) times less than the total maximum. The remaining \( r - 1 \) rows for such a submatrix are chosen from the maximal volume submatrix.

If \( c \) is located in the maximum volume submatrix \( A_M \in \mathbb{C}^{r \times n} \), the result is clear. Otherwise, we combine \( A_M \) with \( c \) and consider the resulting matrix \( A' \in \mathbb{C}^{(r+1) \times n} \). Now, let us apply the lemma 2 and point 1 of the lemma 1 to estimate the ratio of the squared \( r \)-projective
volume of $A'$ to the maximal volume (of $A_M$). In our case, there are $r+1$ possible index sets $T$ for submatrices $A_T \in \mathbb{C}^{r \times n}$ of $A'$.

$$\sum_{|T|=r} V^2(A_T) \left( \frac{V^2(A'_T)}{V^2(A_M)} \right)^2 \geq 1 + \|A^+_M c\|_2^2.$$ 

Here we use the pseudoinverse because $\text{rank } A_M = r$. Next,

$$1 + \|A^+_M c\|_2^2 \geq 1 + \|c\|_2^2 \geq 1 + \|c\|_2^2 \geq 1 + \frac{1}{r},$$

since $c$ is not smaller than any row of $A_M$.

Thus, the sum of the squared volumes of all the submatrices in $A'$ is not less than $1 + 1/r$. By subtracting the volume of $A_M$, the sum of squared volumes of the remaining $r$ submatrices is estimated as at least $1/r$. This implies that the squared volume of some submatrix is at least $1/r^2$. Q. E. D.

Next, we turn to the algorithm RRQR algorithm from [7]. We call this RRQR algorithm GEQR by the first letters of the authors' surnames. Note that the above results improve the number of steps estimate for GEQR to $O\left( r \ln r \ln \frac{m}{r} \right)$ since GEQR also uses the idea of maximizing the submatrix volume.

In practice, the starting submatrix for maxvol is often chosen using the Bebendorf cross algorithm [17]. Despite the fact that for this algorithm the upper bound for elements in $C$ is $2^r$, it is usually enough. The expression for the additional number of steps $k_2$ [5] provides that for $c = O(1)$ even such an estimate on the starting submatrix does not spoil the asymptotics for the number of permutations in maxvol.

Now consider adding rows after $r$-th one from [10]. As in pre-maxvol, the row lengths are the addition criterion.

**Lemma 4 ([10])**. Let $\hat{A}_0 \in \mathbb{C}^{n \times r}$ be a submatrix in the first $r$ rows of the matrix $A \in \mathbb{C}^{M \times r}$. Then adding $i$-th row of the matrix $A$ to the submatrix $\hat{A}_0$ changes the squared volume of $\hat{A}_0$ as

$$\frac{V_2(\hat{A})^2}{V_2(\hat{A}_0)^2} = 1 + l_{0i} = 1 + \|C_{0i:}\|_2^2,$$

where

$$C_0 = A\hat{A}_0^+. $$

The algorithm maxvol2 is presented below. Its conversion formulas are also derived here.

We start with the notation. Like in the lemma, we deal with the matrix $A \in \mathbb{C}^{M \times r}$. Its submatrix $\hat{A} \in \mathbb{C}^{(n+1) \times r}$ expands the submatrix $\hat{A}_0 \in \mathbb{C}^{n \times r}$ by appending a row $a^* = A_{i:}$.

$$\hat{A} = \left[ \begin{array}{c} \hat{A}_0 \\ a^* \end{array} \right].$$

In order to add a row maximizing the volume, we should be able to recalculate the matrix $C_0 \in \mathbb{C}^{M \times n}$ and the squared 2-norms of its rows. They are stored in a vector $l_0 \in \mathbb{C}^M$. For
this we define a matrix \( C \in \mathbb{C}^{m \times n} \) and a column \( C' \in \mathbb{C}^m \) as follows:

\[
A\hat{A}^+ = \begin{bmatrix} C & C' \end{bmatrix}, \quad \hat{A} = \begin{bmatrix} \hat{A}_0 \\ a^* \end{bmatrix}.
\]

They can also be expressed as

\[
C = A(\hat{A}^*\hat{A})^{-1}\hat{A}_0^*, \quad C' = A(\hat{A}^*\hat{A})^{-1}a
\]  \hspace{1cm} (7)

Our task is to calculate \( C \) and \( C' \) on the basis of \( C_0 \). The expression (7) shows that it is sufficient to find \( (\hat{A}^*\hat{A})^{-1} \). Since

\[
\hat{A}^*\hat{A} = \hat{A}_0^*\hat{A}_0 + aa^* = (\hat{A}_0^*\hat{A}_0)\left(1 + (\hat{A}_0^*\hat{A}_0)^{-1}aa^*\right),
\]

then for the inverse

\[
(\hat{A}^*\hat{A})^{-1} = (\hat{A}_0^*\hat{A}_0)^{-1}\left(1 + (\hat{A}_0^*\hat{A}_0)^{-1}aa^*\right)^{-1}
= (\hat{A}_0^*\hat{A}_0)^{-1}\left(1 - \frac{(\hat{A}_0^*\hat{A}_0)^{-1}aa^*}{1 + a^*(\hat{A}_0^*\hat{A}_0)^{-1}a}\right).
\]  \hspace{1cm} (8)

The expression in the brackets can be simplified by introducing the notation \( c^* = C_{i;i}^\dagger \). Indeed,

\[
c^* = C_{0i;i} = A_{i;i}^*\hat{A}_0^* = a^*(\hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*,
\]

\[
c^*c = a^*(\hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\hat{A}_0(\hat{A}_0^*\hat{A}_0)^{-1}a = a^*(\hat{A}_0^*\hat{A}_0)^{-1}a.
\]

Also,

\[
(\hat{A}_0^*\hat{A}_0)^{-1}aa^* = aa^*(\hat{A}_0^*\hat{A}_0)^{-1}.
\]

Substituting into (8) gives

\[
(\hat{A}^*\hat{A})^{-1} = (\hat{A}_0^*\hat{A}_0)^{-1}\left(1 - \frac{aa^*(\hat{A}_0^*\hat{A}_0)^{-1}}{1 + c^*c}\right).
\]

Taking into account

\[
C_0c = A(\hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\hat{A}_0(\hat{A}_0^*\hat{A}_0)^{-1}a = A(\hat{A}_0^*\hat{A}_0)^{-1}a,
\]

this allows to calculate \( C' \) from (7):

\[
C' = A(\hat{A}_0^*\hat{A}_0)^{-1}\left(1 - \frac{a^*a(\hat{A}_0^*\hat{A}_0)^{-1}}{1 + c^*c}\right)a
= A(\hat{A}_0^*\hat{A}_0)^{-1}\left(a - \frac{ac^*c}{1 + c^*c}\right)a
= A(\hat{A}_0^*\hat{A}_0)^{-1}\frac{a}{1 + c^*c}
= \frac{C_0c}{1 + c^*c}
= C_0C_{0i;i}^\dagger/(1 + li).
\]
We compute $C$ similarly:

$$C = A (\hat{A}_0^* \hat{A}_0)^{-1} \left( I - \frac{aa^* (\hat{A}_0^* \hat{A}_0)^{-1}}{1 + c^*c} \right) \hat{A}_0^*$$

$$= C_0 - C'a^* (\hat{A}_0^* \hat{A}_0)^{-1} \hat{A}_0^*$$

$$= C_0 - C'c^*$$

$$= C_0 - C'C_{0i}.$$ 

Now we can directly calculate the lengths of the new rows. The updated $l_0$ is denoted by $l$.

$$l_i = C_{ij}:C_{ij}^* + C'j'$$

$$= (C_{0j} - C'j'C_{0i} C_{0i})^* + |C'|^2$$

$$= l_{0j} - 2C'j'C_{0j} C_{0i} + |C'|^2 C_{0i} C_{0i} + |C'|^2$$

$$= l_{0j} - 2C'j'C_{0j} (1 + l_{0i}) + |C'|^2 l_{0i} + |C'|^2$$

$$= l_{0j} - |C'_j|^2 (1 + l_{0i})$$

Algorithm maxvol2(n) aka maxvol_rect [10]

**Input:** Matrix $A \in \mathbb{C}^{M \times r}$, starting set of row indices $\mathcal{I}$ of cardinality $r$, the required final size $n$.

**Output:** $\mathcal{I}$, supplemented by $n - r$ row indices chosen greedily to maximize the 2-volume.

```plaintext
1: \text{C} := A[A_{\mathcal{I},:}^{-1} 0_{r \times (n-r)}]
2: \text{current\_order} = \{1, \ldots, M\}
3: \text{C.swap}(\mathcal{I},\{1,\ldots,r\},\text{current\_order})
4: l := 0
5: \text{for i := r + 1 to M do}
6: \hspace{1em} l_i := \|C_{i,:}\|_2^2
7: \text{end for}
8: \text{for new\_size := r + 1 to n do}
9: \hspace{1em} l_i := \text{arg max} \|l_i\|_C
10: \hspace{1em} l'_i := 1 + l_i
11: \hspace{1em} l_i := 0
12: C'_i := C_{i,:}/l'_i
13: C' := C'C'_i
14: C := C - C'C_{i,:}
15: \text{for j := new\_size to M do}
16: \hspace{1em} l_j := l_j - l'_i|C'_i|^2
17: \text{end for}
18: C_{j,new\_size} := C'
19: \text{C.swap(new\_size, i, current\_order)}
20: l.swap(new\_size, i)
```

Figure 4: Extension of $\hat{A} = A_{\mathcal{I},:}$ by maxvol2.
This algorithm extends the submatrix by greedily chosen rows to maximize the 2-volume. The transposed version can be applied to select columns in some rows $R$.

Also, maxvol2 provides some guarantees on the ratio of the resulting submatrix volume to the maximum.

**Proposition 2.** If the algorithm maxvol2 is applied to a submatrix that is different in volume from the maximum volume submatrix by no more than $c^{r/2}$ times, where $c \geq 1$ (for example, after the algorithm maxvol with the parameter $c$), then after applying maxvol2 the ratio of the 2-volume of the found submatrix to the 2-volume of the maximum volume submatrix does not exceed

$$V_2/V_1 \leq \max \left( \left( \frac{n}{\tau} \right)^{r/2}, (c^2 \tau)^{r/2} \right). \quad (11)$$

**Proof.** The volume of any submatrix with $n$ rows does not exceed $c^r n^{r/2}$ times the initial matrix volume. Let $x \geq 1/\sqrt{\tau}$. If the ratio of the current volume to the final matrix volume is equal to $x^r n^{r/2}$, at each step the ratio decreases by no less than

$$\sqrt{1 + x^2 \tau} \geq x \sqrt{\tau}$$
times, since it is possible to take rows from the maximum volume submatrix. Hence

$$x^r n^{r/2} \leq \frac{c^r n^{r/2}}{(x \sqrt{\tau})^{n-r}},$$

$$c^r \geq x^n r^{n-r}.$$

If the ratio of volumes turns out to be larger than $c^r n^{r/2}$, then

$$x^r n^{r/2} > c^r n^{r/2} \geq x^n n^{r/2},$$

$$x^{n-r} < n^{r/2} / r^{n/2},$$

$$x n^{1/2} < \frac{n}{\tau} \cdot \left( \frac{n}{\tau} \right)^{1/2} \leq \frac{n}{\tau},$$

which proves (11). $\square$

The algorithm requires $O(Mn(n - r))$ operations to add $n - r$ rows. $C$ matrix can be passed from the algorithm maxvol or pre-maxvol.

Finally, we move on to the Dominant-$C$ algorithm. It is based on the maxvol2 algorithm from above, but now we change the rows instead of adding them. This is done to ensure that the resulting algorithm has properties similar to the maxvol algorithm. Namely, we look for a dominant rectangular submatrix. Its 2-volume does not increase after replacing any single row or column. For the case $n \geq r$, this allows us to ensure that the 2-norm of any row of $C$ outside the rows of the selected submatrix does not exceed $\sqrt{n-r+1}$. It also proves point 2 of lemma 1 and the corollary 1.

The replacement criterion is given in the following lemma.
Lemma 5. Let $\hat{A}_0 \in \mathbb{C}^{n \times r}$ be a submatrix in the first $n$ rows of the matrix $A \in \mathbb{C}^{M \times r}$. Then, replacing $j$-th row of the submatrix $\hat{A}_0$ with the $i$-th row of the matrix $A$ ($i > n$) changes the squared volume of $\hat{A}_0$ by

$$B_{ij} = |C_{0ij}|^2 + (1 + l_{0i})(1 - l_{0j}),$$

times, where the matrix $C_0$ and the vector $l_0$ are the same as in the lemma 4.

Proof. To prove the lemma, we first need to derive the fast recalculation formulas, for which we use recalculation formulas from maxvol2.

Recalculation is done in 2 stages. First, we add the $i$-th row, and then remove $j$-th row. Since we already know how to add, let us only focus on the removal.

Firstly, we note that when the $i$-th row is added, all its elements decrease by $1 + l_{0i}$ times. Which is easily checked by substituting $i$ into the conversion formula for $C$ (line 14 in the algorithm maxvol2). Thus, removing the $j$-th row leads to an increase in its elements by $1 + \tilde{l}_j$ times, where the wave denotes the value after the removal.

$$\tilde{C}_{ij} = C_{ij}(1 + \tilde{l}_j) \quad (12)$$

From the formula for recalculating $l$, we get that

$$l_j = \tilde{l}_j - |C_j'|^2(1 + \tilde{l}_j)
= \tilde{l}_j - |C_{j_0}C_{j_0}'|^2/(1 + \tilde{l}_j)
= \tilde{l}_j - \tilde{l}_j^2/(1 + \tilde{l}_j)
= \frac{\tilde{l}_j}{1 + \tilde{l}_j},$$

which means

$$1 + \tilde{l}_j = 1/(1 - l_j). \quad (13)$$

Using (12) and (13) we obtain a formula for the inverse recalculation of $C$:

$$C := \tilde{C} - C'\tilde{C}_{j_0} = \tilde{C} - C'\tilde{C}_{j_0}(1 + \tilde{l}_j)
= C + C'\tilde{C}_{j_0}/(1 - l_j)$$

Similarly, we obtain the formula for the inverse recalculation of $l$.

$$l_k := \tilde{l}_k - |C_k'|^2(1 + \tilde{l}_j)
= l_k + |C_k'|^2/(1 - l_j)$$

When adding $i$-th row, $l_j$ is connected to $l_{0j}$ as follows:

$$l_j = l_{0j} - |C_j'|^2(1 + l_{0i}) = l_{0j} - |C_{0i}C_{0i}'|^2/(1 + l_{0i}). \quad (14)$$

The product $C_{ij}C_{j_0}'$ is written in the element $C_{ij}$. Indeed,

$$C_{0i}C_{0j} = A_{i_0}(\hat{A}_0')^*\hat{A}_0 = A_{i_0}(\hat{A}_0^*\hat{A}_0)^{-1} = A_{i_0}\hat{A}_0^* = C_{0ij}.$$
Therefore, the addition of the $i$-th row changes the $j$-th row length as

$$l_j = l_{0j} - |C_{0ij}|^2/(1 + l_{0i}). \quad (15)$$

With the help of recalculation formula, we can calculate the ratio of the volumes after the interchange. From the lemma 4 we know that adding the $i$-th row increases the volume by $1 + l_{0i}$ times. The expression (13) shows that removing the $j$-th row results in the volume decrease by $1/(1 - l_j)$ times. The total change considering (15) is equal to

$$B_{ij} := (1 + l_{0i})(1 - l_j) = |C_{0ij}|^2 + (1 + l_{0i}) (1 - l_{0j}).$$

The use of the matrix $B_{ij}$ for the decision to replace the columns leads us to the next algorithm.

**Algorithm Dominant-C($n$)**

**Input:** Matrix $A \in \mathbb{C}^{M \times r}$, starting set of row indices $I$ of cardinality $n$. For example, $I = \{1, \ldots, n\}$.

**Output:** Set of row indices $I$, containing a dominant in 2-volume submatrix.

1: $C := AA^*_I$.
2: for $i := 1$ to $M$ do
3: \hspace{1em} $l_i = \|C_{i,:}\|_2^2 + 1$
4: end for
5: for $i := n + 1$ to $M$ do
6: \hspace{1em} for $j := 1$ to $n$ do
7: \hspace{2em} $B_{ij} := |C_{i,j}|^2 + l_i(1 - C_{j,j})$
8: \hspace{1em} end for
9: end for
10: $B_{i,j} := \arg \max \|B\|_C$
11: while $B_{i,j} > 1$ do
12: \hspace{1em} Recalculate $C$, $l$ and $B$
13: \hspace{1em} Change $j$ to $i$ in $I$
14: \hspace{1em} $B_{i,j} := \arg \max \|B\|_C$
15: end while

The complexity of the algorithm is $O(Mn \cdot \text{iter})$.

Let's estimate the number of iterations after pre-maxvol. We start with a submatrix of volume not more than $\Gamma = (rn)^{r/2}$ times different from the maximum. The use of maxvol2 does not improve the asymptotics, since we take logarithm from $\Gamma$, like in maxvol algorithm (5).

Let the maximum length of a row in $C$ be $x$. The same is true for $C' = CU_{\hat{A}}$. Next, we work with $C'$, since it has only $r$ columns.
Then the volume of the extended submatrix is $1 + x^2$ times greater. The sum of 2-volumes of all its submatrices of size $n \times r$ is $(n - r + 1)(1 + x^2)$. Then among them, there is surely a submatrix with a volume not less than

$$\frac{(n - r + 1)(1 + x^2) - 1}{n} \geq x^2 \frac{n - r + 1}{n}.$$  

Thus, if the initial volume ratio of the current submatrix to the maximal volume submatrix without common rows is equal to $V = \left(\frac{\alpha^2 n^r}{r(n-r+1)}\right)^{r/2}$, then with each step this ratio decreases in $\alpha$ times. So, until $\alpha = c$ (the calculation is the same as for maxvol)

$$k_2 \leq \left\lceil r \ln \frac{\ln \frac{n^r}{r(n-r+1)}}{\ln c} \right\rceil \leq \left\lceil r \ln \frac{2 \ln n}{\ln c} \right\rceil$$

steps are required.

After that, the volume is reduced by $c$ times at each step. So, no more than

$$k_1 \leq \left\lceil r \ln \frac{c^2 \frac{n^r}{r(n-r+1)}}{2 \ln c} \right\rceil \leq \left\lceil r \left(1 + \frac{\ln n}{2 \ln c}\right) \right\rceil - 1$$

additional steps are required. Summarizing and taking into account that only one rounding is enough, we obtain

$$k \leq r \left(1 + \frac{\ln n}{2 \ln c}\right) + r \ln \left(\frac{2 \ln r}{\ln c}\right).$$

If $c = \text{const} > 1$, then

$$k = O(r \ln n).$$

In addition to being a part of maxvol-rect algorithm, this algorithm is interesting in itself. Firstly, it allows to obtain guaranteed estimates of the accuracy of cross approximations in the 2-norm (if used in the first $r$ left and right singular vectors, see Theorem 2 of [12]) and the C-norm (Theorem 4.7 of [15]), which is not guaranteed by the maxvol2 algorithm even when some approximation of rank $r$ is known. Moreover, the estimation of the number of steps allows the algorithm Dominant-C to compete with many algorithms from [16] in the problem of choosing a submatrix with a small norm of the pseudoinverse. The comparison is shown at the end of this section.

In order to find the dominant rectangular submatrix, it remains only to give the algorithm Dominant-R. Now we need to maximize the volume by replacing one of the $r$ columns. Fortunately, such an algorithm already exists. It is the already mentioned GEQR. Dominant-R is a modification of this algorithm, which does not change the asymptotic behavior but allows to reduce the main coefficient due to the specifics of the matrix size ($N \gg n$).

Here the approximation rank is fixed, and only column interchanges are performed. New columns can be added by pre-maxvol.

Lemma 6 ([7], lemma 3.1). Let

$$M = Q \begin{bmatrix} A & B \\ C & \end{bmatrix} \in \mathbb{R}^{m \times N},$$
where QA is a QR decomposition of the submatrix \( \hat{A} \in \mathbb{R}^{m \times r} \) of the matrix M. Then, replacing the j-th column of the submatrix \( \hat{A} \) by the i-th column of the matrix M changes the squared volume of \( \hat{A} \) by

\[
m_{ij} = |(A^{-1}B)_{ij}|^2 + \gamma_i \omega_j
\]

times. Here \( \gamma \) is the vector of the squared 2-norms of the columns of \( C^T \), and \( \omega \) is the vector of the squared 2-norms of the rows of \( A^{-1} \).

Algorithm Dominant-R(r) (The full version is a modification of the column replacement in RRQR from [7])

**Input:** Matrix \( M \in \mathbb{R}^{m \times N} \), starting set of column indices \( I \) of cardinality \( r \), threshold \( f \geq 1 \). For example, \( I = \{1, \ldots, r\} \).

**Output:** Set of column indices \( I \), containing a dominant in 2-volume submatrix.

1. \( Q, A := QR(M\_I) \)
2. \( Q := Q;_{1:r} \)
3. for \( i := 1 \) to \( N - r \) do
4. \( \gamma_i := \|M_{:,i+r}\|^2_2 - \|(Q^T M)_{:,i+r}\|^2_2 \)
5. end for
6. for \( i := 1 \) to \( r \) do
7. \( \omega_i := \|A^{-1}_{i,:}\|^2_2 \)
8. end for
9. \( A^{-1}B := A^{-1} \cdot B \)
10. \( m := \omega^T \gamma \)
11. for \( i := 1 \) to \( r \) do
12. for \( j := 1 \) to \( N - r \) do
13. \( m_{i,j} := m_{i,j} + |(A^{-1}B)_{i+r,j+r}|^2 \)
14. end for
15. end for
16. \( m_{i1,j1} := \arg \max ||m||_C \)
17. while \( m_{i1,j1} > f \) do
18. Change i1 to j1 in I
19. Recalculate Q and A
20. Recalculate \( \gamma \), \( \omega \) and \( A^{-1}B \)
21. \( m := \omega^T \gamma \)
22. for \( i := 1 \) to \( r \) do
23. for \( j := 1 \) to \( N - r \) do
24. \( m_{i,j} := m_{i,j} + |(A^{-1}B)_{i+r,j+r}|^2 \)
25. end for
26. end for
27. \( m_{i1,j1} := \arg \max ||m||_C \)
As in GEQR, replacement of rows can be combined with the addition.

Finally, we combine the algorithms Dominant-C and Dominant-R into a form similar to that for maxvol. This allows us to search for the dominant rectangular submatrix in the entire matrix.

Algorithm maxvol-rect(r, n)

Input: Matrix $A \in \mathbb{R}^{M \times N}$, starting sets of row indices $\mathcal{I}$ and column indices $\mathcal{J}$ of cardinality $n$ and $r$ respectively.

Output: In $\mathcal{I}$ and $\mathcal{J}$ the indices of the rows and columns of the dominant rectangular submatrix of rank $r$ are written.

1: {In the beginning pre-maxvol can be applied.}
2: changed := true
3: old_changed := true
4: while changed do
5: changed := false
6: for changes_in in $\mathcal{I}, \mathcal{J}$ do
7: if changes_in $\in \mathcal{I}$ then
8: $C := A_{\mathcal{I}, \mathcal{J}}$
9: changed := Dominant-C($C, \mathcal{I}$) {Here a small addition to the algorithm is needed, which returns whether there were exchanges}
10: old_changed := changed
11: else
12: $R := A_{\mathcal{I}, \mathcal{J}}$
13: changed := Dominant-R($R, \mathcal{J}$)
14: end if
15: if (not changed) and (changes_in $\in \mathcal{J}$) or (not old_changed) and (changes_in $\in \mathcal{I}$) then
16: changed = false
17: break
18: end if
19: end for
20: end while

Figure 7: Change of the current submatrix $\hat{A} = A_{\mathcal{I}, \mathcal{J}}$ when the algorithm maxvol-rect is executed.
2.1 Improvements and simplifications

Algorithms Dominant-C and Dominant-R can work several times longer than maxvol. Firstly, because of \( n \) instead of \( r \) in the asymptotics. Secondly, because of a larger constant multiplier in front of it. The following is an algorithm that finds the large projective volume submatrix without Dominant-R. It can be used either after maxvol-proj or independently.

**Algorithm maxvol-proj without Dominant-R**

**Input:** Matrix \( A \in \mathbb{R}^{M \times N} \), starting sets of row indices \( I \) and column indices \( J \) of cardinality \( r \) and final sizes of \( C \) and \( R \) (\( n \) and \( m \), respectively).

**Output:** The set \( I \) of the row indices of \( R \) and the set \( J \) of the column indices of \( C \) that contain a submatrix of a large projective volume.

```plaintext
1: changed := true
2: old_changed := true
3: while changed do
4:   changed := false
5:   for changes_in in \{I, J\} do
6:     U, S, V := SVD(A_{I,J})
7:     if changes_in = I then
8:       C := A_{I,J}(V_{1,...,r})^T
9:       changed := Dominant-C(C, I) {Here a small addition to the algorithm is needed, which returns whether there were exchanges}
10:      old_changed := changed
11:   else
12:     R := (U_{i,1,...,r})^T A_{I,J}
13:     changed := Dominant-C(R^T, J)
14:   end if
15:   if ((not changed) and (changes_in = J)) or ((not old_changed) and (changes_in = I)) then
16:     changed = false
17:     break
18:   end if
19: end for
20: end while
```

Multiplication of the columns by the first \( r \) right singular vectors of \( \tilde{A} \) reduces the problem to the search for the large \( r \)-projective submatrix in \( C\tilde{A}_r^+ \). At the same time, we guarantee that the projective volume increases. Indeed, if

\[
\nu_2^2(\tilde{A}^T) > \nu_2^2(\tilde{A}) = 1,
\]

then since

\[
\nu_2^2(\tilde{A}) \nu_2^2(\tilde{A}_r^+) \geq \nu_2^2(\tilde{A}\tilde{A}_r^+),
\]

we get that

\[
\nu_2^2(\tilde{A}) > 1/\nu_2^2(\tilde{A}_r^+) = \nu_2^2(\tilde{A}).
\]
The situation is similar for rows. For $m = r$ or $n = r$, this algorithm is a simplified version of \textit{maxvol-rect}. In this case the row replacements occur only under the condition $| (A^{-1}B)_{ij} | > 1$, and not under the condition $m_{ij} > 1$, which is satisfied more often.

Another version of the projective volume search used in \cite{12} is based on the application of the algorithm \textit{maxvol2}.

**Fast CGR approximation search (2 maxvol2)**

**Input:** Matrix $A \in \mathbb{C}^{M \times N}$, starting sets of row indices $\mathcal{I}$ and column indices $\mathcal{J}$ of cardinality $r$, final sizes of $C$ and $R$ ($n$ and $m$ correspondingly).

**Output:** The set $\mathcal{I}$ of row indices in $R$ and the set $\mathcal{J}$ of column indices in $C$, containing a submatrix with the large projective volume.

1. \textit{maxvol}$(A, \mathcal{I}, \mathcal{J})$
2. $C := A_{\mathcal{I}, \mathcal{J}}$
3. $R := A_{\mathcal{I}, \mathcal{J}}$
4. \textit{maxvol2}$(C, \mathcal{I}, n)$
5. \textit{maxvol2}$(R^T, \mathcal{J}, m)$

Combining all the estimates together with the low probability of starting from a submatrix of small volume, we obtain the overall complexity

\[
\approx O \left( (M + N)r^2 \cdot IT + (M + N)r \cdot \text{iter}_{\text{maxvol}} + (M + N)(m + n)^2 \right).
\]

If instead of searching for dominant submatrices we put a condition with $c = \text{const} > 1$ (although even for $c = 1$ the number of the row interchanges in \textit{maxvol} is of order $r$), then the estimate becomes

\[
\approx O \left( (M + N)r^2 \ln r \cdot IT + (M + N)(m + n)(m + n) \right).
\]

Despite the lack of advantage in terms of asymptotics, this algorithm can work faster because all the replacements of rows and columns are produced in the algorithm \textit{maxvol}, and not in \textit{Dominant-C} and \textit{Dominant-R}. Also, adding one row or column is cheaper than replacing. Nevertheless, as numerical experiments show, this algorithm gives worse approximation than \textit{maxvol-proj}.
Other accelerations and simplifications include: restrictions on the number of switches between rows and columns (high approximation accuracy and large volume are achieved already in step 4), the choice of the coefficient $c > 1$, thereby guaranteeing a small number of replacements (although in practice $c = 1$ works fast enough), a simpler criterion for substitutions (replacing $B_{ij} > 1$ with $|C_{ij}| > 1$ or $(1 + l_i)(1 - l_j) > 1$, which makes it possible not to calculate the matrix $B$) and limiting the number of replacements directly (for example, by the number $n$, which also does not greatly aggravate the accuracy). In addition, the maxvol algorithm can be replaced with simpler analogs. For example, one can be limited to just a greedy set of rows and columns using Householder RRQR with column pivoting or even the Bebendorf algorithm [17]. The use of the latter also allows estimating from above the necessary rank of the approximation in order to achieve the required accuracy.

2.2 Application in the task of subset selection

Algorithms maxvol and Dominant-C can be used to search for the columns containing a strongly nondegenerate submatrix. The obtained estimates of the number of steps allow us to compare them with algorithms from [16]. Point 2 of lemma 1 ensures, that when we select a submatrix $\hat{A} \in \mathbb{C}^{r \times n}$ in the rows $R \in \mathbb{C}^{r \times N}$, we can guarantee the inequality

$$\|\hat{A}^+ R\|_2 \leq \sqrt{1 + \frac{r(N - n)}{n - r + 1}}.$$ 

It immediately follows that

$$\sigma_k(\hat{A}^+) \leq \sigma_k(R^+) \sqrt{1 + \frac{r(N - n)}{n - r + 1}}.$$ 

For the Frobenius norm with the use of one more expression from point 2 of the lemma 1 we obtain

$$\|\hat{A}^+\|_F \leq \|R^+\|_2 \sqrt{1 + \frac{r(N - n)}{n - r + 1}}.$$  (16)

Estimates for maxvol and Dominant-C along with estimates from [16] are given in the table 1. Note that the equivalent to maxvol with the number of columns $k = r$ algorithm GEQR has already been mentioned in [16]. However, we proved a better estimate on the number of steps.

The algorithms are sorted by the number of operations. For Dominant-C, the case
$n = 2r - 1$ and $c = 2$ are given to simplify the comparison. It is assumed that the pre-maxvol algorithm is used up to maxvol or to Dominant-C. Thanks to it we have the estimation of the operation count.

Table 1: Methods for finding a strongly nondegenerate rectangular submatrix $\hat{A} \in \mathbb{C}^{r \times n}$ in the rows $R \in \mathbb{C}^{r \times N}$.

| Method | $\|\hat{A}^+\|_F / \|R^+\|_F^2$ | $\|\hat{A}^+\|_2^2 / \|R^+\|_2^2$ | Complexity |
|--------|---------------------------------|---------------------------------|------------|
| Theorem 3.7 ($\delta = 1/2$) from [16], $n \geq 32k \ln(4k)$ | 4N | 4N | $O(Nr^2 + n \log n)$ |
| maxvol [5], $n = r$ | $(1 + c(N - r)) \frac{r\|R^+\|_2}{\|R^-\|_F}$ | $1 + cr(N - r)$ | $O(Nr^2 \log r / \log c)$ |
| Dominant-C $n = 2r - 1$, $c = 2$ | $(2N + 1) - 4) \frac{r\|R^+\|_2}{\|R^-\|_F}$ | $2(N - 2r) + 3$ | $O(Nr^2 \log r)$ |
| Dominant-C $n \geq r$ | $(1 + \frac{N - n}{n - r + 1}) \frac{r\|R^+\|_2}{\|R^-\|_F}$ | $1 + r\frac{N - n}{n - r + 1}$ | $O(Nnr \log n / \log c)$ |
| Theorem 3.11 ($\delta = 1/2$) from [16], $n = r$ | $c(N - r + 1)$ | $c(N - r + 1)$ | $O(Nr^3 / \log c)$ |
| Theorem 3.5 from [16], $n > r$ | $\frac{(1 + \sqrt{\frac{N}{n}})^2}{(1 - \sqrt{\frac{N}{n}})^2}$ | $\frac{(1 + \sqrt{\frac{N}{n}})^2}{(1 - \sqrt{\frac{N}{n}})^2}$ | $O(Nnr^2)$ |
| Theorem 3.1 from [16], $n \geq r$ | $\frac{N - r + 1}{n - r + 1}$ | $\frac{N - r + 1}{n - r + 1}$ | $O(Nr^2 + N(N - n)r)$ |
| Cons. 3.3 from [16], $n \geq r$ | $\frac{N - r + 1}{n - r + 1} \cdot \frac{r\|R^+\|_2}{\|R^-\|_F}$ | $1 + r\frac{N - n}{n - r + 1}$ | $O(Nr^2 + N(N - n)r)$ |

The last two algorithms in the table are based on the removal of columns. They have a quadratic dependence on $N$, and therefore are applicable only for matrices of small sizes.

According to the estimates for the found submatrices maxvol and Dominant-C algorithms lose only to $r$ times longer working algorithms. With respect to the Frobenius norm, they have worse guarantees only in the case when the ratio $r\frac{\|R^+\|_2}{\|R^-\|_F}$ is not close to $1$. In addition, in the case when the submatrix found by maxvol or Dominant-C has almost equal singular numbers, the factor $r$ also disappears in the estimate for the 2-norm if the expression [16] is used. The only algorithm from [16], capable of competing with Dominant-C in speed, leads to a larger error and requires significantly more columns, while the estimate on the number of operations is only $O(\log r)$ times smaller.

3 Numerical experiments

The question of the effectiveness of the application of the above algorithms for the con-
struction of cross approximations is considered here.

In the case of the cross approximation construction \( n \geq r \) columns \( C \in \mathbb{C}^{M \times n} \) and \( m \geq r \) rows \( R \in \mathbb{C}^{m \times N} \) are selected in the matrix \( A \), and the approximation itself has the form \( CGR \), with *approximation generator* \( G \in \mathbb{C}^{n \times m} \), having rank not higher than \( r \) (for understandable reasons, the term \( CGR \) approximation is often used for these methods).

The error is generally estimated in the 2-norm and the Frobenius norm. The existing lower bounds \([13]\) do not allow to guarantee the high accuracy of approximation in the 2-norm. Nevertheless, recently the result was obtained in the Frobenius norm with the same coefficient as for the \( C \)-norm in \([12]\), but unfortunately, not correct for the maximum volume submatrices.

**Theorem 1 ([11]).** For any matrix \( A \in \mathbb{C}^{M \times N} \) there exists a skeleton approximation \( C\hat{A}^{-1}R \) such that

\[
\|A - C\hat{A}^{-1}R\|_F \leq (r + 1) \|A - A_r\|_F
\]

We hope that this estimate and its analog for the projective volume are often achieved with maximum volume and projective volume search.

So, we expect the ratio of the approximation error to the Frobenius norm error of the best approximation to be the same as for the case of the \( C \)-norm in \([12]\), that is,

\[
E_{U,V}\|A - C\hat{A}^+_rR\|_F \leq \sqrt{\frac{(n + 1)(m + 1)}{(n - r + 1)(m - r + 1)}} \|A - A_r\|_F
\]

in the same notation as before.

In numerical experiments we consider 3 cases: \( m = n = r \), \( m = 2r \), \( n = r \) and \( m = n = 2r \). The corresponding errors, according to our hypothesis, are estimated as follows:

\[
E_{U,V}\|A - C\hat{A}^{-1}R\|_F \leq (r + 1) \|A - A_r\|_F,
E_{U,V}\|A - C\hat{A}^+_rR\|_F \leq \sqrt{2r + 1} \|A - A_r\|_F,
E_{U,V}\|A - C\hat{A}^+_rR\|_F \leq \left(1 + \frac{r}{r + 1}\right) \|A - A_r\|_F.
\]

As we shall see, these estimates are rather crude. However, it is possible to construct an estimate that is much closer to reality, but less grounded.
As point 2 of lemma 1 shows, coefficients of the form $\sqrt{\frac{n+1}{n-r+1}}$ appear from the estimate of the Frobenius norm pseudoinverse of some submatrix $\hat{U} \in \mathbb{C}^{n \times r}$. For the 2-norm estimates, the 2-norm of the pseudoinverse is used. Taking equations from point 2 of the lemma 1 as a replacement for coefficients of the form $\sqrt{\frac{n+1}{n-r+1}}$, we obtain the following estimate:

$$E_{U,V}\|A - C\hat{A}^+R\|_F \approx \sqrt{\left(1 + \frac{1}{M-r}(\|\hat{U}\|_F^2 - r)\right) \left(1 + \left(\frac{1}{N-r}\|\hat{V}\|_F^2 - r\right)\right)}\|F\|_F. \quad (18)$$

Here, $U$ and $V$ are left and right singular vectors of $A$ and $\hat{U}$ and $\hat{V}$ are dominant submatrices of some unitary matrices, not connected with $U$ or $V$. Their sizes for each of the three cases correspond to the sizes of $\hat{A}$. For the matrix sizes $M = N = 100$, the values of the coefficients are shown in the figure 9.

Figure 9: The coefficient in the Frobenius norm of the error in the expression (18) is shown for various sizes of the submatrix $\hat{A}$ (lines with crosses) and the corresponding upper bounds (lines without crosses). The size of the unitary matrix used in the calculation, $100 \times r$, the graph is based on averaging over 100 random matrix generations.
As one can see, the difference is significant. In the following graphs, there are no upper bound to avoid cluttering the figures.

Let’s compare our estimates with the real state of affairs. In the figures 10 and 11 submatrices are constructed through the procedures maxvol [5], maxvol-rect and maxvol-proj.

![Graph](image)

Figure 10: The ratio of the cross methods approximation error to the best Frobenius norm approximation error rate as a function of rank for submatrices of various sizes (lines without crosses) and evaluation of the expression (18) (lines with crosses). Matrix size 100 × 100, the graph is constructed on the basis of averaging over 100 random matrix generations. The k-th largest singular number of the matrix is equal to $1/2^k$.

It can be seen that the upper bound is never violated, and the estimate (18) is reasonably accurate. We note that the relative error is less when the singular numbers of the best error are equal. The approximation is accurate even though in both cases the value of
Figure 11: The ratio of the cross methods approximation error to the best Frobenius norm approximation error rate as a function of rank for submatrices of various sizes (lines without crosses) and evaluation of the expression (18) (lines with crosses). Matrix size $100 \times 100$, the graph is constructed on the basis of averaging over 100 random matrix generations. The first $r$ singular numbers are equal to 10, others to 1.

The Frobenius norm of the approximation error is comparable with the minimum singular number of the approximation.

When the ratio of singular numbers is not 1/10, but 1/1000, the graphs visually coincide entirely.

The figures 12 and 13 show histograms of the error distribution. It is clear that maxvol-proj not only has less error on average, but the error variance is also less.

The figures 14 and 15 show the same histograms for the 2-norm approximation. We see that in the case when the truncated singular value decomposition error in the 2-norm
Figure 12: Distribution of the Frobenius norm error of the approximation by means of various cross algorithms is presented. Matrix size $100 \times 100$, rank of the approximation differs, $\text{maxvol-rect}$ used $2r$ rows. The distribution is based on 1000 random matrix generations. The $k$-th largest singular number of the matrix is equal to $1/2^k$.

differs little from the error in the Frobenius norm, the CGR approximation also has a small error. However, for a large difference of the Frobenius norm from the 2-norm, the cross approximation is much less accurate.

It is also worth noting that the approximation error with maximization of the $r$-projective volume does not depend too much on the maximization method. As already mentioned, in place of $\text{maxvol-proj}$, one can quite quickly apply $\text{maxvol2}$ twice to the found submatrix of size $r \times r$.

In addition, two more methods were considered that work asymptotically longer than
those already cited:

1. maxvol of rank $2r$: The maxvol algorithm for the size $2r \times 2r$ was used. The resulting submatrix was $r$-pseudoinversed.

2. maxvol2r: instead of the second maxvol2, the matrix was expanded by the column $c$ with the maximum value of $\| \hat{A}^+c \|_2$ (justification of efficiency can be found in [12]). The asymptotics of the number of operations is $n$ times worse than for other methods.

Note also that instead of the exact singular value decomposition of the submatrix, an approximate one (for example, RRQR) can be used. Often (but not always!) the error due to this does not grow fast.

As the graph [16] shows, maxvol-proj yields both the smallest mean error and the smallest variance.

The C-norm of the error is studied in detail in [12]. Figure [17] shows that the distribution of the C-norm of the error differs little from the distribution for the Frobenius norm.

Now let us consider how the algorithm approximates non-random matrices. As an example, consider a matrix with elements

$$ (A_1)_{i,j} = (i^{1/3} + j^{1/3})^2 \sqrt{\frac{1}{i} + \frac{1}{j}} \quad (19) $$

which is the so-called “ballistic core”, encountered in the problems of coagulation and fragmentation [14]. The table [2] considers several different sizes. The approximation is sought with an accuracy of $10^{-5} - 10^{-6}$. maxvol-rect, as always, chooses twice the number of rows, compared to the rank. The starting columns and rows are selected randomly.

The table shows that although the cross approximation algorithms in this particular case give results worse than for random matrices, the difference from the singular value decomposition is not too large. Of course, in this case, maxvol-proj should not be used. When the singular numbers decrease so rapidly, it suffices to seek an approximation of rank $n \approx r$ (in the last column, $r + 2$ was used). Then the singular value decomposition for the resulting approximation can be applied if the rank $r$ approximation is required. This technique is
Table 2: The errors of the cross approximation of the matrix $A_1$.

| Size and rank | SVD  | maxvol | maxvol-rect | maxvol-proj | TSVD maxvol of rank $r + 2$ |
|---------------|------|--------|-------------|-------------|-----------------------------|
| $n = 800, r = 12$ | $1.01 \cdot 10^{-5}$ | $2.40 \cdot 10^{-5}$ | $5.73 \cdot 10^{-5}$ | $3.23 \cdot 10^{-5}$ | $1.42 \cdot 10^{-5}$ |
| $n = 400, r = 11$ | $6.09 \cdot 10^{-6}$ | $2.64 \cdot 10^{-6}$ | $2.75 \cdot 10^{-6}$ | $1.59 \cdot 10^{-6}$ | $6.13 \cdot 10^{-6}$ |
| $n = 200, r = 10$ | $3.59 \cdot 10^{-6}$ | $1.23 \cdot 10^{-6}$ | $1.04 \cdot 10^{-6}$ | $7.09 \cdot 10^{-6}$ | $3.59 \cdot 10^{-6}$ |
| $n = 100, r = 9$ | $2.01 \cdot 10^{-6}$ | $5.41 \cdot 10^{-6}$ | $4.87 \cdot 10^{-6}$ | $3.35 \cdot 10^{-6}$ | $2.01 \cdot 10^{-6}$ |

indicated as TSVD (Truncated SVD) for maxvol. In particular, for a given matrix and with $n = r + k$, the first $k$ significant digits of the approximation error coincide with the singular value decomposition error, since the singular values fall approximately as $10^{-k}$.

Nevertheless, in the case of a close distribution of the singular numbers truncated SVD of the constructed approximation no longer gives any advantages. On the other hand, maxvol-proj becomes more efficient. For another example, we manually replace the singular numbers of the error of $A_1$ with the equal ones. That is, we consider a matrix $A_2$ such that

$$A_2 = U_1 \Sigma_2 V_1,$$

where

$$A_1 = U_1 \Sigma_1 V_1$$

is a singular value decomposition of $A_1$ and

$$\sigma_i(\Sigma_2) = \sigma_i(\Sigma_1), \quad i = 1, \ldots, r,$$

$$\sigma_i(\Sigma_2) = \sigma_j(\Sigma_2), \quad i, j > r,$$

$$\|(A_1) - (A_1)_r\|_F = \|(A_2) - (A_2)_r\|_F.$$ 

Thus, SVD gives the same error. The table 3 shows the results of the approximation of the matrix $A_2$ by the same algorithms.

Table 3: The errors of the cross approximation of the matrix $A_2$.

| Size and rank | SVD  | maxvol | maxvol-rect | maxvol-proj | TSVD maxvol of rank $2r$ |
|---------------|------|--------|-------------|-------------|-----------------------------|
| $n = 800, r = 12$ | $1.01 \cdot 10^{-5}$ | $2.02 \cdot 10^{-5}$ | $1.71 \cdot 10^{-5}$ | $1.44 \cdot 10^{-5}$ | $1.59 \cdot 10^{-5}$ |
| $n = 400, r = 11$ | $6.09 \cdot 10^{-6}$ | $1.19 \cdot 10^{-6}$ | $9.63 \cdot 10^{-6}$ | $9.01 \cdot 10^{-6}$ | $9.94 \cdot 10^{-6}$ |
| $n = 200, r = 10$ | $3.59 \cdot 10^{-6}$ | $6.86 \cdot 10^{-6}$ | $6.03 \cdot 10^{-6}$ | $5.03 \cdot 10^{-6}$ | $5.57 \cdot 10^{-6}$ |
| $n = 100, r = 9$ | $2.01 \cdot 10^{-6}$ | $3.84 \cdot 10^{-6}$ | $3.30 \cdot 10^{-6}$ | $2.71 \cdot 10^{-6}$ | $3.11 \cdot 10^{-6}$ |

The table shows that in this case, maxvol-proj gives the best result. Truncated SVD with
maxvol of higher rank (here in the last column the rank \( n = 2r \) is used) is much less useful than for \( A_1 \). Note that in this case simple maxvol also performs better. However, its error still grows with size and rank. For maxvol-proj the error ratio never exceeds 1.5 times the error of SVD. It is also worth noting that the use of Truncated SVD still improves accuracy: the direct application of maxvol of rank \( 2r \) gives a larger error. Thus, in the case of the too slow decay of the singular numbers, the use of maxvol of higher rank may not yield any advantages. To increase the accuracy the rank of the approximation must be reduced. Naturally, one can also reduce the rank after maxvol-proj, if the slow decrease in singular numbers is found. Such a decrease can be verified in the constructed approximation.

4 Conclusion

In general, low-rank approximations based on the maximum volume principle require very few operations and show high accuracy of the approximation. When a choice between the speed and accuracy of the approximation is needed, one can use the algorithm maxvol-proj itself, including its version without Dominant-R, or its simplifications. At the same time, the necessary rank can be estimated from above, starting with the Bebendorf algorithm [17]. If necessary, one can reduce it with the help of Truncated SVD applied to the constructed cross approximation. In the case of a rapid decrease of singular numbers, Truncated SVD can be applied to an approximation of a larger rank and lead to a significant improvement in accuracy in comparison to the direct search for an approximation of rank \( r \). This makes it possible in some cases to achieve better accuracy of the approximation without maxvol-proj. It should be noted however that maxvol-proj is a universal method in the sense that the accuracy of the cross approximation based on it in the Frobenius norm does not depend (on average) on the distribution of the singular numbers of the matrix.

Numerical experiments showed that averaging over random matrices of left and right singular vectors gives the same form of the error coefficient in the Frobenius norm as in the \( C \)-norm. It turns out to be directly related to the estimates for the Frobenius norm of pseudoinverse to the submatrices of unitary matrices. Moreover, the estimates of this coefficient from above coincide with the known lower bounds both for \( n = r \) and for \( n \to \infty \) [3, 9].
5 Appendix

A Full versions of algorithms

Here detailed descriptions of the algorithms and brief derivations for some new recalculation formulas are presented. The justifications of the recalculation formulas for maxvol and GEQR are in the corresponding articles. References are given in the names of the algorithms. Recalculation for DominantC is provided in the Algorithms section, lemma [5]

The author of this article is of the opinion that algorithms should be written so that their writing in any programming language is as close as possible to simple copying.

Algorithm maxvol(r) [5]

Input: Matrix $A \in \mathbb{C}^{M \times N}$, starting sets of row indices $I$ and column indices $J$ of cardinality $r$. For example, $I = J = \{1, \ldots, r\}$.

Output: In $I$ and $J$ the indices of the rows and columns of the rank-$r$ dominant submatrix are written.

1: $C := A_{i,j} A_{i,j}^{-1}$
2: current_order = $\{1, \ldots, M\}$
3: C.swap($I, \{1, \ldots, r\}$, current_order) {C.swap(A, B, order)
swaps the elements of C by rearranging the sets of A and B indexes and correspondingly changing the order in order. }
4: $C_{i,j} := \arg \max \|C\|_C$
5: changed := true
6: old_changed := true
7: while changed do
8: changed := false
9: for changes_in in $\{I, J\}$ do
10: {Here, the pointer is assigned, and no copies of the sets are made. }
11: while $|C_{i,j}| > 1$ do
12: if $i = j$ then
13: break
14: end if
15: C := $C - C_{i,j} (C_{i,:} - e_j) / C_{i,j}$
16: C.swap(i, j, current_order)
17: $C_{i,j} := \arg \max \|C\|_C$
18: changed := true
old_changed := true
end while
if ((not changed) and (changes_in = J))
or ((not old_changed) and (changes_in = I)) then
changed = false
break
end if
changes_in := current_order[1..r] \{The numbering starts at 1.\}
if changes_in = J then
old_changed := false
C := A_{\mathcal{I}, j} A_{\mathcal{J}, \mathcal{J}}^{-1}
current_order = \{1, \ldots, M\}
C.swap(\mathcal{I}, \{1, \ldots, r\}, current_order)
else
C := A_{\mathcal{I}, i} A_{\mathcal{I}, \mathcal{J}}^{-T}
current_order = \{1, \ldots, N\}
C.swap(\mathcal{J}, \{1, \ldots, r\}, current_order)
end if
C_{i,j} := \arg \max_{C} \|C\|_C
end for
end while

\textbf{Algorithm Dominant-C(n)}
\textbf{Input:} Matrix $A \in \mathbb{C}^{M \times r}$, starting set of row indices $\mathcal{I}$ of cardinality $n$. For example, $\mathcal{I} = \{1, \ldots, n\}$.
\textbf{Output:} Set of row indices $\mathcal{I}$, containing a dominant in 2-volume submatrix.

1: $C := AA_{\mathcal{I}, \mathcal{I}}^+$
2: current_order := \{1, \ldots, M\}
3: C.swap(\mathcal{I}, \{1, \ldots, n\}, current_order)
4: $l := 0_M$
5: for $i := 1$ to $M$ do
6: $l_i = \|C_{i,}:\|_2$
7: end for
8: $B := 0_{M \times n}$
9: for $i := n + 1$ to $M$ do
10: for $j := 1$ to $n$ do
11: $B_{ij} := |C_{ij}|^2 + (1 + l_i)(1 - l_j)$
12: end for
13: end for
14: $B_{i,j} := \arg \max_{B} \|B\|_C$
while $B_{i,j} > 1$ do

16: $C'_i := C_{i,C}/(1 + l_i)$
17: $C' := CC'_i$
18: for $k := 1$ to $M$ do
19: $l'_k := l_k - |C'_k|^2/(1 + l_i)$
20: end for
21: $C := C - C'C_i$:
22: $C'.swap(i,j) , current_order)$
23: $C'.swap(i,j)$
24: $l'.swap(i,j)$
25: swap($C,i,j$,current_order)
26: for $k := 1$ to $M$ do
27: $l_k := l'_k + |C'_k|^2/(1 - l'_i)$
28: end for
29: $C := C + C'(C_{i,C} / (1 - l'_i))$
30: for $i := n + 1$ to $M$ do
31: for $j := 1$ to $n$ do
32: $B_{i,j} := |C_{i,j}|^2 + (1 + l_i)(1 - l_j)$
33: end for
34: end for
35: $B_{i,j} := \arg \max \|B\|_C$
36: end while
37: $I := current_order[1..n]$

Algorithm Dominant-R($r$) (modified RRQR from [7])

As in [7] the matrix sizes are denoted by $m$ and $n$. The rank of approximation is $k$. In our case $n \gg m$, while in [7] the case $m \geq n$ is considered. This allows us not to worry about performance when recalculating the $Q$ matrix and not to store it as a product of reflections.

Some important changes in notation:

- $\gamma$ (vector, there is also a scalar $\gamma$) contains squared lengths of the columns of $C^T$.
- $\omega$ contains squared lengths of the rows of $A^{-1}$.

Input: Matrix $M \in \mathbb{R}^{m \times N}$, starting set of column indices $I$ of cardinality $r$, threshold $f \geq 1$.

For example, $I = \{1, ..., r\}$.

Output: Set of column indices $I$, containing a dominant in 2-volume submatrix.

1: current_order := \{1, ..., n\}
2: $M\_swap\_columns(I,\{1, ..., k\}, current\_order)$
   \{the same as swap, but for the columns.\}
3: $Q, A := QR(M_I)$
4: \( Q := Q_{i,1:k+1} \)
5: \( \gamma := 0_{n-k} \)
6: for \( i := 1 \) to \( n - k \) do
7: \( \gamma_i := \|M_{i,i+k}\|_2^2 \)
8: end for
9: \( B := Q_{i,k}^T M \)
10: for \( i := 1 \) to \( n - k \) do
11: \( \gamma_i := \gamma_i - \|B_{i,i+k}\|_2^2 \)
12: end for
13: \( \omega := 0_k \)
14: for \( i := 1 \) to \( k \) do
15: \( \omega_i := \|A_{i-1,i}\|_2^2 \)
16: end for
17: \( A^{-1}B := A^{-1} \cdot B \)
18: \( m := \omega^T \gamma \)
19: for \( i := 1 \) to \( k \) do
20: for \( j := 1 \) to \( n - k \) do
21: \( m_{i,j} := m_{i,j} + \|A^{-1}B\|_{i+k,j+k}^2 \)
22: end for
23: end for
24: \( m_{i1,j1} := \arg \max \|m\|_C \)
25: \( \rho := m_{i1,j1} \)
26: while \( \rho > f \) do

\{Exchange of \( i_1 \) and \( k \)\}
27: for \( i := i_1 \) to \( k - 1 \) do
28: \( \omega \).swap\((i, i + 1)\)
29: \((A^{-1}B)\).swap_columns\((i, i + 1)\)
30: \(A\).swap\((i, i + 1)\)
31: \(M\).swap_columns\((i, i + 1, \text{current_order})\)
32: end for
33: for \( i := i_1 \) to \( k - 1 \) do
34: \( \alpha, \beta := \text{givens}(A_{i,i}, A_{i+1,i}) \) \{see [6], page 240\}
35: \( A_{i:i+1,k} := \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} A_{i:i+1,k} \)
36: \( Q_{i,i+1} := Q_{i,i+1} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \)
37: end for

\{Exchange \( k, k + 1, i_1 \Rightarrow j_1, k, k + 1 \} \}
b_c := M_{:,j1}
M.swap_columns(j1,k+1,current_order)
M.swap_columns(k,k+1,current_order)
γ.swap(j1,k+1)
(A^{-1}B).swap_columns(j1,k+1)
c_1^T := A_{k,k}(A^{-1}B)_{k,k+2:n}
γν := \sqrt{c_{k+1}}
γ := A_{k,k}
b_1 := A_{1:k-1,k}
A_{:,k} := A(A^{-1}B)_{:,k+1}
Q_{:,k+1} := (b_c - Q_{:,1:k}A_{:,k})/(γν)
m := (b_1^T - b_c^TQ_{:,1:k}Q_{:,1:k}^T)/(γν)
c_2^T := mM_{:,k+2:n}

\{Nullification of k-th column under the diagonal\}
α, β := givens(A_{k,k}, γν)
Q_{:,i:i+1} := Q_{:,i:i+1} \begin{bmatrix} α & β \\ -β & α \end{bmatrix}
\begin{bmatrix} γμ/ρ & \bar{c}_1^T \\ γν/ρ & \bar{c}_2^T \end{bmatrix} := \begin{bmatrix} α & -β \\ β & α \end{bmatrix} \begin{bmatrix} γ & c_1^T \\ 0 & c_2^T \end{bmatrix}
A_{k,k} := \sqrt{(γν)^2 + A_{k,k}^2}

\{Recalculation of γ\}
γ_{k+1} := (γν/ρ)^2
for i := k + 2 to n do
γ_i := γ_i + |\bar{c}_1^T_i-k-1|^2 - |\bar{c}_2^T_i-k-1|^2
end for

\{Recalculation of ω\}
u := A_{k-1:k-1}.RTsolve(b_1) \{Solving the system with the upper triangular matrix\}
ω_k := 1/A_{k,k}^2
μ := (A^{-1}B)_{k,k+1}
μ̃ := (A^{-1}B)_{1:rank-1,rank+1} + μu
for i := 1 to k - 1 do
ω_i := ω_i + ω_k|μ̃_i|^2 - |μ_i/γ|^2
end for

\{Recalculation of A^{-1}B\}
This is a modification of GEQR without recalculation of the matrices $B$ and $C^T$. This reduces in 2 times the coefficient of $mn$ in the complexity estimate of a single iteration.

All designations and recalculations are the same as in the original article [7]. They are not repeated here. We focus only on the main changes: removing the recalculation of $B$ and $C$.

We got rid of the $B$ recalculation quite simply: note that we only need the first row and the last column of $B$. And they are easily computed through $A$ and $A^{-1}B$ (lines 43 and 47).

As for the $C^T$ matrix, we only need to know the first row. Since unitary transformations were performed in GEQR for the matrix $C^T$, the first element of this row is equal to the 2-norm of the first column (line 44). All other elements are obtained as scalar products with the corresponding normalized column of the matrix $B$ in the subspace orthogonal to the first $k$ columns of $M$ (lines 49-50).

To rotate, we still need to know $k + 1$-st column of the matrix $Q$ (line 48). It is easily found from the representation of the column...
\( b_c \) of the matrix \( M \) as a product of \( Q \) by the last column of the extension of the matrix \( A \):

\[
b_c = \begin{bmatrix} Q_{:,1:k} & Q_{:,k+1} \\ A_{:,k} & \gamma \end{bmatrix}.
\]

Compared to GEQR, we thus save one product of vectors of size \( m - k \) and \( n - k \) and one addition with a matrix of size \((m - k) \times (n - k)\). When calculating \( C \), there was also a multiplication of the matrix of the specified size by a vector. In Dominant-R we multiply the matrix of size \( m \times (n - k) \) by a vector.

We can apply the same tactics to add the columns. We do not repeat the derivation of the changes, the description of the original algorithm is in [7]. Here the benefit is even greater (of course, provided that \( A^{-1}B \) is needed later) since at each iteration the width of the submatrix is less than \( k \).

As before, the case of \( k = 1 \) is not considered separately, although many operations are absent for it. The maximum \( k \) is denoted by \( r \).

**Algorithm pre-maxvol(\( r \)) (modification of the column addition in RRQR from [7])**

**Input:** Matrix \( M \in \mathbb{R}^{m \times n} \), required rank \( r \).

**Output:** Set of column indices \( \mathcal{I} \) of size \( r \), containing a submatrix, whose volume differs from the maximum by no more than \( r! \) times.

1: \( \mathcal{I} := \emptyset \)
2: \( \omega := 0_r \)
3: \( A := 0_{r \times r} \)
4: \( A^{-1}B := 0_{r \times n} \)
5: \( Q := I_m \)
6: \( \gamma := 0_n \)
7: for \( i := 1 \) to \( n \) do
8: \( \gamma_i := \|M_i\|_2^2 \)
9: end for
10: \( \gamma_j := \text{arg max} \|\gamma\|_C \)
11: for \( k := 1 \) to \( r \) do
12: \( \mathcal{I} := \mathcal{I} \cup \{j\} \)

\{Exchange of \( j \) and \( k \}\)
13: \( b_c := M_{:,j} \)
\begin{verbatim}
M.swap_columns(j, k, current_order)
γ.swap(j, k)
(A^{-1}B).swap_columns(j, k)
γν := √c_k
A_{:,k-1} := A_{:1:k-1,1:k-1}(A^{-1}B)_{1:k-1,k}
A_{k,1} := γν
Q_{:,k} := (b_c - Q_{:,1:k}A_{:,k})/(γν)
m := (b_c^T - b_c^TQ_{:,1:k}Q_{:,1:k-1}^T)/(γν)
c_2^T := mM_{:,k+1:n}

{Recalculation of γ}
γ_k := 0
for i := k+1 to n do
γ_i := γ_i - |c_{i:k-1}^T|^2
end for

{Recalculation of ω}
ω_k := 1/A_{k,k}^2
for i := 1 to k-1 do
ω_i := ω_i + ω_k|A^{-1}B|_{i,k}^2
end for

{Recalculation of A^{-1}B}
(A^{-1}B)_{k,k+1:n} := c_2^T/A_{k,k}
(A^{-1}B)_{1:1:k-1,k+1:n} := (A^{-1}B)_{1:1:k-1,k+1:n} - (A^{-1}B)_{1:1:k-1,k+1}(A^{-1}B)_{k,k+1:n}
γ_j := arg max ||γ||_C
end for

References

[1] Goreinov S.A., Tyrtyshnikov E.E. The maximal-volume concept in approximation by low-rank matrices // Contemporary Mathematics. — 2001. — Vol. 268. — P. 47–51.

[2] Goreinov S.A., Tyrtyshnikov E.E., Zamarashkin N.L. A theory of pseudo-skeleton approximations // Linear Algebra and Its Applications. — 1997. — Vol. 261. — P. 1–21.
\end{verbatim}
[3] A. Deshpande, L. Rademacher et al. Matrix Approximation and Projective Clustering via Volume Sampling // Theory of Computing. — 2006. — Vol. 2. — P. 225–247.

[4] Goreinov S.A., Tyrtyshnikov E.E. Quasioptimality of skeleton approximation of a matrix in the Chebyshev norm // Doklady Mathematics. — 2011. — Vol. 83, no. 3. — P. 1–2.

[5] How to find a good submatrix / S.A. Goreinov, I.V. Oseledets, D.V. Savostyanov et al. // Matrix Methods: Theory, Algorithms, Applications / Ed by V. Olshevsky, E. Tyrtyshnikov. — World Scientific Publishing, 2010. — P. 247-256.

[6] G. H. Golub and C. F. Van Loan. Matrix Computations. — 4th ed. — The Johns Hopkins University Press, Baltimore, MD, 2013.

[7] M. Gu, S. C. Eisenstat, efficient algorithms for computing a strong rank-revealing qr factorization // SIAM J. Sci. COMPUT. — 1996. — Vol. 17, no. 4. — P. 848-869.

[8] A. I. Osinsky. Probabilistic estimation of the rank 1 cross approximation accuracy // arXiv 1706.10285 (Submitted on 30 Jun 2017).

[9] A. Deshpande and S. Vempala. Adaptive sampling and fast low-rank matrix approximation // Approximation, randomization and combinatorial optimization — 2006. — Vol. 4110 of LNCS, Springer, Berlin. — P. 292–303.

[10] Mikhalev A.Y., Oseledets I.V. Rectangular submatrices of maximum volume and their computation // Doklady Mathematics — 2015. — Vol. 91, no. 3. — P. 267-268.

[11] Zamarashkin N.L., Osinsky A.I. On the Existence of a Nearly Optimal Skeleton Approximation of a Matrix in the Frobenius Norm // Doklady Mathematics — 2018. — Vol. 97, no. 2. — P. 164-166.
[12] А.И. Осinsky, Н.Л. Замарашкин. Псевдо-
skeleton approximations with better accu-
accuracy estimates // Linear Algebra and its Ap-
lications — 2018. — Vol. 537. — P. 221-249.

[13] C. Boutsidis, P. Drineas, M. Magdon-Ismail.
Near-optimal column-based matrix recon-
struction // SIAM Journal on Computing —
2013.

[14] Быстрый метод решения уравнений
агрегационно-фрагментационной
кинетики типа уравнений Смолуховского
/ С. А. Матвеев, Е. Е. Тыртышников, А.
П. Смирнов, Н. В. Бриллиантов // Вычислительные методы и программирование:
Новые вычислительные технологии (Электронный научный журнал). — 2014. — Т. 15,
№ 1. — С. 1–8.

[15] Michalev A.Y., Oseledets I.V. Rectangular
maximum-volume submatrices and their ap-
plications // Linear Algebra and its Applica-
tions. — 2018. — Vol. 538. — P. 187–211.

[16] H. Avron and C. Boutsidis. Faster Subset
Selection for Matrices and Applications //
SIAM J. Matrix Anal. Appl. — 2011. — Vol.
34, no. 4.

[17] M. Bebendorf. Approximation of bound-
ary element matrices // Numerische Math-
ematic. — 2000. — Vol. 86. — P. 565–589.

[18] Ali Çivril, Malik Magdon-Ismail. On select-
ing a maximum volume sub-matrix of a ma-
trix and related problems // Theoretical
Computer Science. — 2009. — Vol. 410, no.
47-49. — P. 4801–4811.
Figure 13: Distribution of the Frobenius norm error of the approximation by means of various cross algorithms is presented. Matrix size $100 \times 100$, rank of the approximation differs, maxvol-rect used $2r$ rows. The distribution is based on 1000 random matrix generations. The singular numbers of the error are 1000 times smaller than the singular numbers of the best approximation.
Figure 14: Distribution of the 2-norm error of the approximation by means of various cross algorithms. Matrix size $100 \times 100$, rank of the approximation differs, maxvol-rect used $2r$ rows. The distribution is based on 1000 random matrix generations. The $k$-th largest singular number of the matrix is equal to $1/2^k$. 
Figure 15: Distribution of the 2-norm error of the approximation by means of various cross algorithms is presented. Matrix size $100 \times 100$, rank of the approximation differs, maxvol-rect used $2r$ rows. The distribution is based on 1000 random matrix generations. The singular numbers of the error are 1000 times smaller than the singular numbers of the best approximation.
Figure 16: Distribution of the error in the Frobenius norm of the cross approximation by means of a submatrix of a large $r$-projective volume maximized by various methods. Matrix size $100 \times 100$, rank of approximation 10. Submatrix size $20 \times 20$. The distribution is based on 2000 random matrix generations. The first 10 singular numbers are equal to 10, the remaining to 1.
Figure 17: Distribution of the C-norm error for the cross approximation by means of various cross methods. Matrix size $100 \times 100$, rank of approximation 10. Submatrix size $20 \times 20$. The distribution is based on 2000 random matrix generations. The first 10 singular numbers are equal to 10, the remaining to 1.