Low Rank Approximation of a Matrix at Sub-linear Cost

Victor Y. Pan

Department of Computer Science
Lehman College of the City University of New York
Bronx, NY 10468 USA and
Ph.D. Programs in Mathematics and Computer Science
The Graduate Center of the City University of New York
New York, NY 10036 USA
victor.pan@lehman.cuny.edu
http://comet.lehman.cuny.edu/vpan/

Abstract

A matrix algorithm performs at *sub-linear cost* if it uses much fewer flops and memory cells than the input matrix has entries. Using such algorithms is indispensable for Big Data Mining and Analysis, where the input matrices are so immense that one can only access a small fraction of all their entries. Typically, however, such matrices admit their Low Rank Approximation (LRA), which one can access and process at sub-linear arithmetic cost, that is, by involving much fewer memory cells and arithmetic operations than an input matrix has entries. Can, however, we compute LRA at sub-linear cost? Adversary argument shows that no algorithm running at sub-linear cost can output accurate LRA of the worst case input matrices, or even of the matrices of small families of our Appendix A, but for more than a decade Cross–Approximation (CA) iterations, running at sub-linear cost, have routinely been computing accurate LRA. We partly resolve that long-known contradiction by proving that already a single two-stage C–A loop computes reasonably close LRA of any matrix close to a matrix of sufficiently low rank provided that the C-A loop begins at a submatrix that shares its numerical rank with an input matrix. We cannot obtain such an initial submatrix for the worst case input matrix without accessing all or most of its entries, but have this luck with a high probability for any choice from a random input matrix and increase our chances for success with every new C-A iteration. All this should explain the well-known empirical power of C-A iterations applied to real world inputs.

Keywords: Cross–Approximation, CUR Low Rank Approximation, Maximal volume

2000 Math. Subject Classification: 65Y20, 65F30, 68Q25, 15A52

1 Introduction

1.1 LRA problem and our main result

Low rank approximation (LRA) of a matrix[1] is a fundamental subject of Numerical Linear Algebra NLA) and Computer Science (CS). An $m \times n$ matrix W admits its close approxi-

[1] Here and hereafter the concepts “low”, “large”, “small”, “far”, “close”, etc. are defined in context.

mation of rank at most r if and only if the matrix W has *numerical rank* at most r (then we write $\text{nrank}(W) \leq r$) or equivalently if and only if

$$W = AB + E, \ ||E||/||W|| \leq \epsilon, \quad (1)$$

for $A \in \mathbb{C}^{m\times r}$, $B \in \mathbb{C}^{r\times n}$, a small integer r, a matrix norm $||\cdot||$, and a small tolerance ϵ. Such an LRA approximates the mn entries of W by using $(m+n)r$ entries of A and B rather than the mn entries of the matrix W, which is dramatic saving in the case where $(m+n)r \ll mn$. (Here and hereafter inequalities $a \ll b$ and $b \gg a$ show that the ratio $|a/b|$ is small in context.) This is a crucial benefit in applications of LRA to Big Data Mining and Analysis, where the input matrices are so immense that one can only access a tiny fraction of all their entries, and one can further save memory space by representing LRA, in its special form of CUR LRA (see Section 2).

Can we, however, compute close LRA *at sub-linear cost*, that is, by using much fewer flops and memory cells than the input matrix has entries? Based on adversary argument one can prove that no algorithm running at sub-linear cost can output close LRA of the worst case inputs and even of the matrices of a small families of our Appendix A, but for more than a decade Cross–Approximation (CA) iterations, running at sub-linear cost, have been routinely computing close CUR LRA worldwide. Our main result (see Corollary 23, Remark 24, and the end of Appendix A) provides long-missing partial formal support for this empirical phenomenon.

Namely we restrict the class of input matrices to those close to matrices of sufficiently low rank and assume that a loop of two C–A iterations begins at a submatrix that shares its numerical rank with an input matrix. Then we readily prove that this loop outputs a reasonably close CUR LRA.

We cannot choose an initial submatrix sharing numerical rank with a worst case input matrix without accessing all or most of its entries of input matrix and similarly for the small families of input matrices in Appendix A), but with a high probability this property holds for any submatrix of a random input matrix [PLSZa], which can already explain empirical power of C-A iterations dealing with real world input matrices. Moreover it is sufficient to obtain such a desired submatrix for any C-A iteration, and the chances only increase at the next C-A iteration together with the volume of the submatrix.

Our study provides new insight into LRA by means of C-A iterations. Our upper estimates for the output errors of LRA slowly increase as an input matrix deviates from a low rank matrix, but that restriction is rather mild (see Appendix C). These upper estimates grow exponentially in numerical rank of an input matrix, thus restricting our main result to inputs having a small constant numerical rank, but we firmly believe that with more work that restrictions can be relaxed.

1.2 Earlier works

The reader can access extensive bibliography on LRA and CUR LRA via [HMT11], [M11], [W14], [CBSW14], [OZ16], [KS16], [BW17], [SWZ17], [OZ18], [P18], and the references therein.

The study of CUR (aka CGR and pseudo-skeleton) LRA can be traced back to the skeleton decomposition in [G59] and QRP factorization in [G65] and [BG65], redefined and refined as rank-revealing factorization in [C87].

The CUR LRA algorithms in [CH90], [CH92], [HP92], [HLY92], [CI94], [GE96], and [P00] largely rely on the maximization of the volume $(\det(G^*G))^{1/2}$ of a CUR generator G

(which is a submatrix of an input matrix). This fundamental idea goes back to [K85] and has been developed in [GZT95], [T96], [GTZ97], [GTZ97a], [GT01], [GOSTZ10], [GT11], [M14], and most recently in [OZ18].

The study in [GZT95], [T96], [GTZ97], and [GTZ97a] towards volume maximization revealed the crucial property that the computation of LRA requires no factorization of the input matrix but just proper selection of its row and column sets.

C–A iterations were a natural extension of this observation preceded by the Alternating Least Squares method of [CC70] and [H70] and leading to dramatic empirical decrease of quadratic memory space and cubic arithmetic time used by LRA algorithms. The concept of C–A was implicit in [T96] and coined in [T00]; we credit [B00], [BR03], [GOSTZ10], [OT10], [B11], [KV16], and [KV16] for devising efficient C–A and adaptive C-A algorithms.

Our present results appeared in arxiv reports [PLSZ16, Section 5] and [PLSZ17, Part II] together with various results on LRA of random input matrices.[2]

1.3 Organization of our paper

We define CUR LRA and C–A iterations in the next section. We devote Section 3 to background material on matrix volumes, their maximization and its impact on LRA. In Section 4 we recall C-A iterations and in Section 5 prove that they output reasonably close LRA of a matrix having sufficiently low numerical rank. We recall relevant definitions of matrix computations and some auxiliary results in the Appendix, where we also specify some small families of input matrices that admit close LRA but are hard for computing it at sub-linear cost.

2 Background: CUR LRA

We use basic definitions for matrix computations recalled in Appendix B.

CUR LRA of a matrix W of numerical rank at most r is defined by three matrices C, U, and R, with C and R made up of l columns and k rows of W, respectively, $U \in \mathbb{C}^{l\times k}$ said to be the *nucleus* of CUR LRA,[3]

$$0 < r \le k \le m,\ r \le l \le n,\ kl \ll mn, \tag{2}$$

$$W = CUR + E,\ \text{and}\ ||E||/||W|| \le \epsilon,\ \text{for a small tolerance}\ \epsilon > 0. \tag{3}$$

CUR LRA is a special case of LRA of (1), say, for $A = LU$, $B = R$, and $k = l = r$. Conversely, given LRA of (1) one can compute CUR LRA of (3) at sub-linear cost (see [PLa] and [PLSZa]).

Define a *canonical* CUR LRA as follows.

(i) Fix two sets of columns and rows of W and define its two submatrices C and R made up of these columns and rows, respectively.

[2]The pioneering papers [PLSZ16] and [PLSZ17] provide first formal support for LRA at sub-linear cost, which they call "superfast" LRA. That work, unsuccessfully submitted to ACM STOC 2017, has extended to LRA the earlier techniques of [PQY15], [PZ17a], and [PZ17a], proposed for the analysis of randomized Gaussian elimination with no pivoting and other fundamental matrix computations.

[3]The pioneering papers [GZT95], [GTZ97], [GTZ97a], [GT01], [GT11], [GOSTZ10], [M14], [OZ16], and [OZ18] define CGR approximations having nuclei G; "G" can stand, say, for "germ". We use the acronym CUR, more customary in the West. "U" can stand, say, for "unification factor", and we notice the alternatives of CNR, CCR, or CSR with N, C, and S standing for *"nucleus", "core",* and *"seed"*.

(ii) Define the $k \times l$ submatrix $W_{k,l}$ made up of all common entries of C and R, and call it *CUR generator*.

(iii) Compute its rank-r truncation $W_{k,l,r}$ by setting to 0 all its singular values, except for the r largest ones.

(iv) Compute the Moore–Penrose pseudo inverse $U =: W_{k,l,r}^{+}$ and call it the *nucleus* of CUR LRA of the matrix W (cf. [DMM08], [OZ18]); see alternative choices of nuclei in [MD09], [BW17], [SWZ17].

Notice that $W_{r,r} = W_{r,r,r}$, and if a CUR generator $W_{r,r}$ is nonsingular, then $U = W_{r,r}^{-1}$.

3 Background: Matrix Volumes

3.1 Definitions and the Hadamard's bound

Definition 1. For a triple of integers k, l, and r such that $1 \le r \le \min\{k, l\}$, the *volume* $v_2(M)$ and the *r-projective volume* $v_{2,r}(M)$ of a $k \times l$ matrix M are defined as follows:

$$v_2(M) := \prod_{j=1}^{\min\{k,l\}} \sigma_j(M), \quad v_{2,r}(M) := \prod_{j=1}^{r} \sigma_j(M), \tag{4}$$

$$v_{2,r}(M) = v_2(M) \text{ if } r = \min\{k, l\}, \tag{5}$$

$$v_2^2(M) = \det(MM^*) \text{ if } k \ge l;\ v_2^2(M) = \det(M^*M) \text{ if } k \le l,\ v_2^2(M) = |\det(M)|^2 \text{ if } k = l.$$

By following [CI94], [GZT95], [GTZ97], [GTZ97a], [GE96], [P00], [GT01], [GOSTZ10], [GT11], [M14], [OZ16], and [OZ18], we use the concepts of volume and projectuve volume in our study of CUR LRA; [B-I92] shows some distinct applications of the concept of projective volume.

Definition 2. The volume of a $k \times l$ submatrix $W_{\mathcal{I},\mathcal{J}}$ of a matrix W is *h-maximal* over all $k \times l$ submatrices if it is maximal up to a factor of h. The volume $v_2(W_{\mathcal{I},\mathcal{J}})$ is *column-wise* (resp. *row-wise*) h-maximal if it is h-maximal in the submatrix $W_{\mathcal{I},:}$ (resp. $W_{:,\mathcal{J}}$). The volume of a submatrix $W_{\mathcal{I},\mathcal{J}}$ is *column-wise* (resp. *row-wise*) *locally* h-maximal if it is h-maximal over all submatrices of W that differ from the submatrix $W_{\mathcal{I},\mathcal{J}}$ by a single column (resp. single row). Call volume (h_c, h_r)-*maximal* if it is both column-wise h_c-maximal and row-wise h_r-maximal. Likewise define *locally* (h_c, h_r)-*maximal* volume. Call 1-maximal and $(1, 1)$-maximal volumes *maximal*. Extend all these definitions to r-projective volumes.

For a $k \times l$ matrix $M = (m_{ij})_{i,j=k,l}^{r}$ write $\mathbf{m}_j := (m_{ij})_{i=1}^{k}$ and $\bar{\mathbf{m}}_i := ((m_{ij})_{j=1}^{l})^*$ for all i and j. For $k = l = r$ recall the *Hadamard's bound*

$$v_2(M) = |\det(M)| \le \min\ \{\prod_{j=1}^{r} ||\mathbf{m}_j||,\ \prod_{i=1}^{r} ||\bar{\mathbf{m}}_j^*||,\ r^{r/2} \max_{i,j=1}^{r} |m_{ij}|^r\}. \tag{6}$$

3.2 Volume maximization and bounding a singular value

Hereafter we write

$$t_{q,r,h}^2 := (q - r)rh^2 + 1. \tag{7}$$

Lemma 3. *(See [CI94], [GZT95], [GTZ97], [GTZ97a], [GE96], [P00], [M14], [OZ16], and [OZ18].) Suppose that* $\min\{h, h'\} \geq 1$, $W_{\mathcal{I},\mathcal{J}} \in \mathbb{C}^{k\times l}$ *is a submatrix of a matrix* $W \in \mathbb{C}^{m\times n}$*. Then*

$$t_{n,r,h}\ \sigma_r(W_{\mathcal{I},\mathcal{J}}) \geq \sigma_r(W_{\mathcal{I},:})$$

if $k = r \leq l$ *and if the volume* $v_2(W_{\mathcal{I},\mathcal{J}})$ *is locally column-wise* h*-maximal and*

$$t_{m,r,h'}\ \sigma_r(W_{\mathcal{I},\mathcal{J}}) \geq \sigma_r(W_{:,\mathcal{J}})$$

if $k \geq l = r$ *and if this volume is locally row-wise* h'*-maximal.*

Notice that $v_2(W_{\mathcal{I},\mathcal{J}}) = v_{2,r}(W_{\mathcal{I},\mathcal{J}})$ for the above matrices $W_{\mathcal{I},\mathcal{J}}$ of sizes $r \times l$ and $k \times r$.

Proof. The lemma turns into [P00, Lemma 3.5] for $k = l = r$ and is extended to the case where $r = \min\{k, l\}$ because no singular value of a matrix increases in the transition to its submatrix. □

3.3 The impact of volume maximization on CUR LRA

The estimates of the two following theorems in the Chebyshev matrix norm $||\cdot||_C$ increased by a factor of $\sqrt{mn}$ turn into estimates in the Frobenius norm $||\cdot||_F$ (see (11)).

Theorem 4. *[OZ18].*[4] *Suppose that* $r := \min\{k, l\}$, $W_{\mathcal{I},\mathcal{J}}$ *is the* $k \times l$ *CUR generator,* $U = W^+_{\mathcal{I},\mathcal{J}}$ *is the nucleus defining a canonical CUR LRA of an* $m \times n$ *matrix* W, $E = W - CUR$, $h \geq 1$, *and the volume of* $W_{\mathcal{I},\mathcal{J}}$ *is* h*-maximal, that is,*

$$h\ v_2(W_{\mathcal{I},\mathcal{J}}) = \max_B v_2(B)$$

where the maximum is over all $k \times l$ *submatrices* B *of the matrix* W*. Then*

$$||E||_C \leq h\ f(k, l)\ \sigma_{r+1}(W) \quad \text{for} \quad f(k, l) := \sqrt{\frac{(k+1)(l+1)}{|l-k|+1}}.$$

Theorem 5. *[OZ18]. Suppose that* $W_{k,l} = W_{\mathcal{I},\mathcal{J}}$ *is a* $k \times l$ *submatrix of an* $m \times n$ *matrix* W, $U = W^+_{k,l,r}$ *is the nucleus of a canonical CUR LRA of* W, $E = W - CUR$, $h \geq 1$, *and and the* r*-projective volume of* $W_{\mathcal{I},\mathcal{J}}$ *is* h*-maximal, that is,*

$$h\ v_{2,r}(W_{\mathcal{I},\mathcal{J}}) = \max_B v_{2,r}(B)$$

where the maximum is over all $k \times l$ *submatrices* B *of the matrix* W*. Then*

$$||E||_C \leq h\ f(k, l, r)\ \sigma_{r+1}(W) \quad \text{for} \quad f(k, l, r) := \sqrt{\frac{(k+1)(l+1)}{(k-r+1)(l-r+1)}}.$$

Observe the following corollary of Theorem 28.

Corollary 6. Suppose that $BW = (BU|BV)$ for a nonsingilar matrix B and that the submatrix U is h-maximal in the matrix $W = (U|V)$. Then the submatrix BU is h-maximal in the matrix BW.

[4] The theorem first appeared in [GT01, Corollary 2.3] in the special case where $k = l = r$ and $m = n$.

4 C–A iterations

Next we describe C–A iterations by involving two auxiliary Sub-algorithms $\mathcal{A}$ and $\mathcal{B}$.

For a fixed 4-tuple of integers k, l, p, and q such that $r \le k \le p \le m$ and $r \le l \le q \le n$ Sub-algorithm $\mathcal{A}$ is applied to a $p \times q$ submatrix $\widehat{W}$ of W and computes a $k \times l$ submatrix of $\widehat{W}$ whose volume or projective volume is maximal up to a fixed factor $h \ge 1$ among all $k \times l$ submatrices of $\widehat{W}$.

Sub-algorithm $\mathcal{B}$ verifies whether the error norm of the CUR LRA built on a fixed CUR generator is within a fixed tolerance τ (see [PLa] on some verification recipes).

For simplicity one can first consider the C-A algorithm in the case where $k = l = r$ (see Figure 1, borrowed from [PLSZa]).

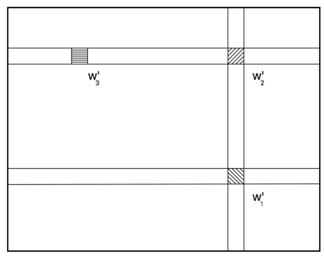

Figure 1: The three successive C–A steps output three striped matrices.

Algorithm 7. C–A iterations.

Input: An $m \times n$ matrix W, a target rank r for its CUR LRA, a target size $k \times l$ of a CUR generator such that (2) holds, a positive τ, and a positive integer ITER.

Output: A CUR LRA of W with error norm at most τ or FAILURE.

Initialization: Fix a submatrix W_0 made up of l columns of W.

Computations: The algorithm recursively alternates "vertical" and "horizontal" C–A steps. The ith C-A step is "vertical" for even i and "horizontal" for odd i.

In both cases at the i-th step Sub-algorithm $\mathcal{A}$ is applied to a $p \times q$ submatrix W_i of W and outputs a $k \times l$ submatrix W_i' of W_i, which is used as a CUR generator in order to build on it a CUR LRA of W.

At the "vertical" steps, $p := m$, $q := l$, and $W_i := W_{:,\mathcal{J}_i}$ is an $m \times l$ matrix made up of l columns of W.

At the "horizontal" steps, $p := k$, $q := n$, and $W_i := W_{\mathcal{I}_i,:}$ is a $k \times n$ matrix made up of k rows of W.

Sub-algorithm $\mathcal{B}$ verifies whether the error norm of the CUR LRA built of the CUR generator $W_i' := W_{\mathcal{I}_i,\mathcal{J}_i}$ is within the tolerance τ. If so, the CUR LRA is output and the computation stops.

If $i \ge$ ITER, the computation stops and FAILURE is output.

Otherwise the $k \times l$ submatrix W_i' of W is embedded into a submatrix W_{i+1} of W of size $k \times n$ for even i and of size $m \times l$ for odd i, and the next $(i+1)$st C–A step is initiated at this submatrix.

5 CUR LRA by Means of C–A Iterations

We can apply C–A steps by choosing deterministic algorithms of [GE96] for Sub-algorithm $\mathcal{A}$. In this case ml and kn memory cells and $O(ml^2)$ and $O(k^2n)$ flops are involved in "vertical" and "horizontal" C-A iterations, respectively. They are superfast if $k^2 = o(m)$ and $l^2 = o(n)$ and output submatrices having h-maximal volumes for h being a low degree polynomial in $m+n$. Every iteration outputs a matrix that has locally h-maximal volume in a "vertical" or "horizontal" submatrix, and the hope is to obtain globally $\bar{h}$-maximal submatrix (for reasonably bounded h) when maximization is performed recursively in alternating directions.

Remark 8. Alternative algorithms of [P00] do the same as those of [GE96], although they square the h of [GE96]. Empirically the algorithms of both [GE96] and [P00] are superseded by the algorithm *maxvol* of [GOSTZ10].

Of course, the contribution of C-A step is nil where it is applied to a $p \times q$ input whose volume 0 or nearly vanishes compared to the target maximum, but the consistent success of C-A iterations in practice suggests that in a small number of loops such a degeneration is regularly avoided.

Next we show that it is avoided and that already two successive C-A iterations output a CUR generator having h-maximal volume and projective volume (and then we estimate h) in the case where the iterations begin at a $p \times q$ submatrix of W that shares its rank $r > 0$ with W. By continuity of the volume the result is extended to small perturbations of such matrices.

In the next two subsections we consider *the worst case input matrix* W of a rank r and two successive C-A steps initiated at its two submatrices of rank r. In this case we prove that the $k \times l$ output matrix $W_{k,l}$ for $\min\{k,l\} = r$ has $\bar{h}$-maximal volume among $k \times l$ submatrices of the input matrix W for $\bar{h}$ of order $(mnr)^{r/2}$, which decreases to order of r^r if $k = l = r$. This order is reasonable for small r, and in that case we arrive at a reasonably close CUR LRA of the matrix W by virtue of Theorem 4.

In Section 5.3 we extend these results to the maximization of r-projective volume rather than the volume of a CUR generator. (Theorem 5 shows benefits of such a maximization.)

In Section 5.4 we summarize our study in this section and comment on the estimated and empirical performance of C–A iterations.

5.1 From locally to globally h-maximal volumes of full rank submatrices

Theorem 9. *Suppose that a $r \times l$ submatrix U has a nonzero column-wise locally h-maximal volume in the matrix $W = (U \mid V) \in \mathbb{C}^{r \times n}$ for $h \geq 1$. Then this submatrix has $\widehat{h}$-maximal volume in the matrix W for $\widehat{h} = t_{n,r,h}^r$ for $t_{q,r,h}$ of (7).*

Proof. By means of orthogonalization of the rows of the matrix W obtain its factorization $W = RQ$ where R is a $r \times r$ nonsingular matrix and $Q = (R^{-1}U \mid R^{-1}V)$ is a $r \times n$ unitary matrix and deduce from Corollary 6 that the volume of the matrix $R^{-1}U$ is column-wise locally h-maximal in the matrix Q.

Therefore $\sigma_r(R^{-1}U) \geq \sigma_r(Q)/t_{n,r,h}$ by virtue of Lemma 3.

Combine this bound with the relationships $\sigma_r(Q) = 1$ and $v_2(R^{-1}V) \geq (\sigma_r(R^{-1}V))^r$ and deduce that $\widehat{h} v_2(R^{-1}U) \geq 1$ for $\widehat{h} = t_{n,r,h}^r$ of equation (7).

Notice that $v_2(Q_l) \leq v_2(Q) = 1$ for any $r \times l$ submatrix Q_l of Q.

Hence the volume $v_2(R^{-1}U)$ is $\widehat{h}$-maximal in Q.

Now Theorem 9 follows from Corollary 6. $\square$

Example 10. The bound of Theorem 9 is quite tight for $r = h = 1$. Indeed the unit row vector $\mathbf{v} = \frac{1}{\sqrt{n}}(1, \dots, 1)^T$ of dimension n is a $r \times n$ matrix for $r = 1$. Its coordinates are $r \times r$ submatrices, all having volume $\frac{1}{\sqrt{n}}$. Now notice that $\sqrt{n} \approx \widehat{h} = ((n-1)+1)^{1/2} = t_{n,1,1}$ (cf. equation (7)).

Remark 11. The theorem is readily extended to the case of a $k \times n$ matrix W of rank r, $0 < r \leq k \leq n$, where r-projective volume replaces volume. Indeed row ortogonalization reduces the extended claim precisely to Theorem 9.

Next we decrease the upper bound $\widehat{h} = t^r_{n,l,h}$ of Theorem 9 in the case where $l = r$ (cf. [GOSTZ10]). We begin with a lemma.

Lemma 12. *Let $W = (I_r \mid V) \in \mathbb{C}^{r \times n}$ for $r \leq n$ and let the submatrix I_r have column-wise locally h-maximal volume in W for $h \geq 1$. Then $||W||_C \leq h$.*

Proof. Let $|w_{ij}| > h$ for an entry w_{ij} of the matrix W, where, say, $i = 1$. Interchange its first and jth columns. Then the leftmost block I_r turns into the matrix $R = \begin{pmatrix} w_{1j} & \mathbf{u}^T \\ \mathbf{0} & I_{r-1} \end{pmatrix}$. Hence $v_2(R) = |\det(R)| = |w_{1j}| > h$. Therefore I_r is not a column-wise locally h-maximal submatrix of W. The contradiction implies that $||W||_C \leq h$. $\square$

Theorem 13. *Suppose that $r \times r$ submatrix U has a nonzero column-wise locally h-maximal volume in a matrix $W = (U \mid V') \in \mathbb{C}^{r \times n}$ for $h \geq 1$. Then this submatrix has $\tilde{h}$-maximal volume in W for $\tilde{h} = h^r r^{r/2}$.*

Proof. Apply Lemma 12 to the matrix $U^{-1}W = (I_r \mid V)$ for $V = U^{-1}V'$ and obtain that $||U^{-1}W||_C \leq h$. Hadamard's bound (6) for $M = V$ implies that the volume 1 of the submatrix I is $\tilde{h}$-maximal in the matrix $U^{-1}W$ for the claimed value of $\tilde{h}$. Now deduce from Corollary 6 that the submatrix U has $\tilde{h}$-maximal volume in W. $\square$

Remark 14. Clearly the bound $\widehat{h} = t^r_{n,r,h}$ of Theorem 9 is larger than the bound $\tilde{h} = r^{r/2}$ of Theorem 13, but how much larger? Substitute a slightly smaller expression $((k-r)rh^2)^{1/2}$ for $t_{k,r,h} = ((k-r)rh^2 + 1)^{1/2}$ of equation (7) into the equation $\widehat{h} = t^r_{n,r,h}$ and observe that the resulting decreased value is still larger than $\tilde{h} = h^r r^{r/2}$ by a factor of $(n-r)^{r/2}$.

5.2 Volume of the output of a C–A loop

First we compare SVDs of two matrices W and W^+ and obtain the following lemma.

Lemma 15. *$\sigma_j(W)\sigma_j(W^+) = 1$ for all matrices W and all subscripts j, $j \leq \mathrm{rank}(W)$.*

Corollary 16. $v_2(W)v_2(W^+) = 1$ and $v_{2,r}(W)v_{2,r}(W^+) = 1$ for all matrices W of full rank and all integers r such that $1 \leq r \leq \mathrm{rank}(W)$.

Now we are ready to prove that for some specific constants g and h nonzero volume of a $k \times l$ submatrix of a rank-r matrix W is g-maximal globally, that is, over all its $k \times l$ submatrices, if it is h-maximal locally, over the $k \times l$ submatrices of two input matrices of two successive C-A steps.

Theorem 17. *Suppose that the volume of a $k \times l$ submatrix $W_{\mathcal{I},\mathcal{J}}$ is nonzero and (h,h')-maximal in a matrix W for $h \geq 1$ and $h' \geq 1$ where* $\mathrm{rank}(W) = r = \min\{k,l\}$. *Then this volume is hh'-maximal over all its $k \times l$ submatrices of the matrix W.*

Proof. The matrix $W_{\mathcal{I},\mathcal{J}}$ has full rank because its volume is nonzero.

Fix any $k \times l$ submatrix $W_{\mathcal{I}',\mathcal{J}'}$ of the matrix W, recall that $W = CUR$, and obtain that

$$W_{\mathcal{I}',\mathcal{J}'} = W_{\mathcal{I}',\mathcal{J}} W^+_{\mathcal{I},\mathcal{J}} W_{\mathcal{I},\mathcal{J}'}.$$

If $k \leq l$, then first apply claim (iii) of Theorem 28 for $G := W_{\mathcal{I}',\mathcal{J}}$ and $H := W^+_{\mathcal{I},\mathcal{J}}$; then apply claim (i) of that theorem for $G := W_{\mathcal{I}',\mathcal{J}} W^+_{\mathcal{I},\mathcal{J}}$ and $H := W_{\mathcal{I},\mathcal{J}'}$ and obtain that

$$v_2(W_{\mathcal{I}',\mathcal{J}} W^+_{\mathcal{I},\mathcal{J}} W_{\mathcal{I},\mathcal{J}'}) \leq v_2(W_{\mathcal{I}',\mathcal{J}}) v_2(W^+_{\mathcal{I},\mathcal{J}}) v_2(W_{\mathcal{I},\mathcal{J}'}).$$

If $k > l$ deduce the same bound by applying the same argument to the matrix equation

$$W^T_{\mathcal{I}',\mathcal{J}'} = W^T_{\mathcal{I},\mathcal{J}'} W^{+T}_{\mathcal{I},\mathcal{J}} W^T_{\mathcal{I}',\mathcal{J}}.$$

Combine this bound with Corollary 16 for W replaced by $W_{\mathcal{I},\mathcal{J}}$ and deduce that

$$v_2(W_{\mathcal{I}',\mathcal{J}'}) = v_2(W_{\mathcal{I}',\mathcal{J}} W^+_{\mathcal{I},\mathcal{J}} W_{\mathcal{I},\mathcal{J}'}) \leq v_2(W_{\mathcal{I}',\mathcal{J}}) v_2(W_{\mathcal{I},\mathcal{J}'}) / v(W_{\mathcal{I},\mathcal{J}}). \quad (8)$$

Recall that the matrix $W_{\mathcal{I},\mathcal{J}}$ is (h,h')-maximal and conclude that

$$h v_2(W_{\mathcal{I},\mathcal{J}}) \geq v_2(W_{\mathcal{I},\mathcal{J}'}) \text{ and } h' v_2(W_{\mathcal{I},\mathcal{J}}) \geq v_2(W_{\mathcal{I}',\mathcal{J}}).$$

Substitute these inequalities into the above bound on the volume $v_2(W_{\mathcal{I}',\mathcal{J}'})$ and obtain that $v_2(W_{\mathcal{I}',\mathcal{J}'}) \leq hh' v_2(W_{\mathcal{I},\mathcal{J}})$. □

5.3 From maximal volume to maximal r-projective volume

Recall that the CUR LRA error bound of Theorem 4 is strengthened when we shift to Theorem 5, that is, maximize r-projective volume for $r < k = l$ rather than the volume. Next we reduce maximization of r-projective volume of a CUR generators to volume maximization.

Corollary 6 implies the following lemma.

Lemma 18. *Let M and N be a pair of $k \times l$ submatrices of a $k \times n$ matrix and let Q be a $k \times k$ unitary matrix. Then $v_2(M)/v_2(N) = v_2(QM)/v_2(QN)$, and if $r \leq \min\{k,l\}$ then also $v_{2,r}(M)/v_{2,r}(N) = v_{2,r}(QM)/v_{2,r}(QN)$.*

Algorithm 19. [*From the maximal volume to the maximal r-projective volume.*]

INPUT: Four integers k, l, n, and r such that $0 < r \leq k$ and $r \leq l \leq n$, a $k \times n$ matrix W of rank r and a black box algorithm that computes a $r \times l$ submatrix of maximal volume in a $r \times n$ matrix of full rank r.

OUTPUT: A column set $\mathcal{J}$ such that the $k \times l$ submatrix $W_{:,\mathcal{J}}$ has maximal r-projective volume in the matrix W.

COMPUTATIONS: 1. Compute a rank-revealing QRP factorization $W = QRP$, where Q is a unitary matrix, P is a permutation matrix, $R = \begin{pmatrix} R' \\ O \end{pmatrix}$, and R' is a $r \times n$ matrix.[5] (See [GL13, Sections 5.4.3 and 5.4.4] and [GE96].)

[5] One can apply other rank-revealing factorizations instead.

2. Compute a $r \times l$ submatrix $R'_{:,\mathcal{J}}$ of R' having maximal volume $v_2(R')$ and output the matrix $W_{:,\mathcal{J}}$.

The submatrices R' and $\begin{pmatrix} R' \\ O \end{pmatrix}$ have maximal volume and maximal r-projective volume in the matrix R, respectively, by virtue of Theorem 28 and because $v_2(\bar{R}) = v_{2,r}(\bar{R}) = v_{2,r}(R')$. Therefore the submatrix $W_{:,\mathcal{J}}$ has maximal r-projective volume in the matrix W by virtue of Lemma 18.

Remark 20. By transposing a horizontal input matrix W and interchanging the integers m with n and k with l we extend the algorithm to computing a $k \times l$ submatrix of maximal or nearly maximal r-projective volume in an $m \times l$ matrix of rank r.

5.4 Complexity and accuracy of a two-step C–A loop

By combining Theorems 9 and 13 deduce that the volume of a $r \times l$ submatrix is $\bar{h}$-maximal in a $r \times n$ matrix of rank r for $\bar{h} = t^r_{n,r,h}$ if $l > r$ and for $\bar{h} = h^r r^{r/2}$ if $l = r$ provided that the volume of the submatrix is column-wise locally h-maximal. In this case we obtain that Algorithm 19 computes a $k \times l$ submatrix having maximal r-projective volume in an $m \times n$ matrix of rank r for any 5-tuple of integers k, l, m, n, and r such that $r \le k \le m$ and $r \le l \le n$. The following theorem summarizes these observations.

Theorem 21. *Given five integers k, l, m, n, and r such that $r \le k \le m$ and $r \le l \le n$, suppose that two successive C–A steps (say, based on the algorithms of [GE96] or [P00]) combined with Algorithm 19 have been applied to an $m \times n$ matrix W of rank r and have output $k \times l$ submatrices W'_1 and $W'_2 = W_{\mathcal{I}_2,\mathcal{J}_2}$ with nonzero r-projective column-wise locally h-maximal and nonzero r-projective row-wise locally h'-maximal volumes, respectively. Then the submatrix W'_2 has $\bar{h}$-maximal r-projective volume in the matrix W for $\bar{h} := (t_{n,r,h} t_{m,r,h'} r)^r$ for $t^2_{q,r,h} = (q-r)rh^2 + 1$ of equation (7). The bound on $\bar{h}$ decreases to $(hh'r)^r$ if $k = l = r$.*

Proof. By applying Algorithm 19 reduce the claim of the theorem to the case where r-projective volume is equal to the volume of a matrix of full rank r. Then combine Theorems 9, 13, and 17. □

Remark 22. [Cf. Remark 14.] How sharp is the estimate $\bar{h} := (t_{n,r,h} t_{m,r,h'} r)^r$ of Theorem 21? Substitute a slightly smaller expression $((k-r)rh^2)^{1/2}$ for $t_{k,r,h} = ((k-r)rh^2+1)^{1/2}$ into the product $(t_{n,r,h} t_{m,r,h'} r)^r$. Then its value decreases but still exceeds the bound $(hh'r)^r$ by a factor of $((m-r)(n-r))^{r/2}$.

In this section we arrived at a C–A algorithm that computes a CUR approximation of a rank-r matrix W. Let us summarize our study by combining Theorems 4, 5, and 21.

Corollary 23. Under the assumptions of Theorem 21 apply a two-step C–A loop to an $m \times n$ matrix W and suppose that both its C–A steps output $k \times l$ submatrices having nonzero r-projective column-wise and row-wise locally h-maximal volumes (see Remarks 24 and 8). Build a canonical CUR LRA on a CUR generator $W'_2 = W_{k,l}$ of rank r output by the second C–A step. Then

(i) the computation of this CUR LRA by using the auxiliary algorithms of [GE96] or [P00] involves $(m+n)r$ memory cells and $O((m+n)r^2)$ flops[6] and

[6]For $r = 1$ an input matrix turns into a vector of dimension m or n, and then we compute its absolutely maximal coordinate just by applying $m-1$ or $n-1$ comparisons, respectively.

(ii) the error matrix E of the output CUR LRA satisfies the bound $||E||_C \le g(k,l,r)\ \bar{h}\ \sigma_{r+1}(W)$ for $\bar{h}$ of Theorem 21 and $g(k,l,r)$ denoting the functions $f(k,l)$ of Theorem 4 or $f(k,l,r)$ of Theorem 5. In particular $||E||_C \le 2hh'\sigma_2(W)$ for $k=l=r=1$.

Remark 24. Theorem 27 enables us to extend Algorithm 19, Theorem 21, and Corollary 23 to the case of an input matrix W of numerical rank r provided that the volume of the $k \times n$ input submatrix of C–A iterations stays nonzero in the transition from this matrix W to its LRA W'.

Appendix

A Small families of hard inputs for sub-linear cost LRA

Any sub-linear cost LRA algorithm fails on the following small families of LRA inputs.

Example 25. Define the following family of $m \times n$ matrices of rank 1 (we call them *δ-matrices*): $\{\Delta_{i,j},\ i=1,\dots,m;\ j=1,\dots,n\}$. Also include the $m \times n$ null matrix $O_{m,n}$ into this family. Now fix any sub-linear cost algorithm; it does not access the (i,j)th entry of its input matrices for some pair of i and j. Therefore it outputs the same approximation of the matrices $\Delta_{i,j}$ and $O_{m,n}$, with an undetected error at least $1/2$. Apply the same argument to the set of $mn+1$ small-norm perturbations of the matrices of the above family and to the $mn+1$ sums of the latter matrices with any fixed $m \times n$ matrix of low rank. Finally, the same argument shows that a posteriori estimation of the output errors of an LRA algorithm applied to the same input families cannot run at sub-linear cost.

This example actually covers randomized LRA algorithms as well. Indeed suppose that an LRA algorithm does not access a constant fraction of the entries of an input matrix. Then with a constant probability the algorithm misses an entry whose value greatly exceeds those of all other entries, in which case the algorithm can hardly approximate that entry closely. We show, however, that close LRA can be computed at sub-linear cost in two successive C-A iterations provided that we avoid choosing degenerating initial submatrix, which is precisely the problem with the matrix families of Example 25. The sub-linear cost algorithms of [MW17] and [BW18] compute LRA of matrices of two important special matrix classes.

B Definitions for matrix computations and a lemma

Next we recall some basic definitions for matrix computations (cf. [ABBB99], [GL13]).

$\mathbb{C}^{m\times n}$ is the class of $m \times n$ matrices with complex entries.

I_s denotes the $s \times s$ identity matrix. $O_{q,s}$ denotes the $q \times s$ matrix filled with zeros.

$\text{diag}(B_1,\dots,B_k) = \text{diag}(B_j)_{j=1}^k$ denotes a $k \times k$ block diagonal matrix with diagonal blocks $B_1,\dots,B_k$.

$(B_1\ |\ \dots\ |\ B_k)$ and $(B_1,\dots,B_k)$ denote a $1 \times k$ block matrix with blocks $B_1,\dots,B_k$.

W^T and W^* denote the transpose and the Hermitian transpose of an $m \times n$ matrix $W = (w_{ij})_{i,j=1}^{m,n}$, respectively. $W^* = W^T$ if the matrix W is real.

For two sets $\mathcal{I} \subseteq \{1,\dots,m\}$ and $\mathcal{J} \subseteq \{1,\dots,n\}$ define the submatrices

$$W_{\mathcal{I},:} := (w_{i,j})_{i\in\mathcal{I};j=1,\dots,n}, W_{:,\mathcal{J}} := (w_{i,j})_{i=1,\dots,m;j\in\mathcal{J}},\ \text{and}\ W_{\mathcal{I},\mathcal{J}} := (w_{i,j})_{i\in\mathcal{I};j\in\mathcal{J}}. \quad (9)$$

An $m\times n$ matrix W is *unitary* (also *orthogonal* when real) if $W^*W=I_n$ or $WW^*=I_m$. *Compact SVD* of a matrix W, hereafter just *SVD*, is defined by the equations

$$W=S_W\Sigma_W T_W^*, \tag{10}$$

$$\text{where } S_W^*S_W=T_W^*T_W=I_\rho,\ \Sigma_W:=\operatorname{diag}(\sigma_j(W))_{j=1}^{\rho},\ \rho=\operatorname{rank}(W),$$

$\sigma_j(W)$ denotes the jth largest singular value of W for $j=1,\dots,\rho$; $\sigma_j(W)=0$ for $j>\rho$.

$||W||=||W||_2$, $||W||_F$, and $||W||_C$ denote spectral, Frobenius, and Chebyshev norms of a matrix W, respectively, such that (see [GL13, Section 2.3.2 and Corollary 2.3.2])

$$||W||=\sigma_1(W),\ ||W||_F^2:=\sum_{i,j=1}^{m,n}|w_{ij}|^2=\sum_{j=1}^{\operatorname{rank}(W)}\sigma_j^2(W),\ ||W||_C:=\max_{i,j=1}^{m,n}|w_{ij}|,$$

$$||W||_C\le ||W||\le ||W||_F\le \sqrt{mn}\ ||W||_C,\ ||W||_F^2\le \min\{m,n\}\ ||W||^2. \tag{11}$$

$W^+:=T_W\Sigma_W^{-1}S_W^*$ is the Moore–Penrose pseudo inverse of an $m\times n$ matrix W.

$$||W^+||\sigma_r(W)=1 \tag{12}$$

for a full rank matrix W.

A matrix W has *ϵ-rank* at most $r>0$ for a fixed tolerance $\epsilon>0$ if there is a matrix W' of rank r such that $||W'-W||/||W||\le\epsilon$. We write $\operatorname{nrank}(W)=r$ and say that a matrix W has *numerical rank* r if it has ϵ-rank r for a small ϵ.

Lemma 26. *Let $G\in\mathbb{C}^{k\times r}$, $\Sigma\in\mathbb{C}^{r\times r}$ and $H\in\mathbb{C}^{r\times l}$ and let the matrices G, H and Σ have full rank $r\le\min\{k,l\}$. Then $||(G\Sigma H)^+||\le ||G^+||\ ||\Sigma^+||\ ||H^+||$.*

Proof. For the sake of completeness we include a proof of this well-known result.

Let $G=S_G\Sigma_G T_G$ and $H=S_H\Sigma_H T_H$ be SVDs where S_G, T_G, D_H, and T_H are unitary matrices, Σ_G and Σ_H are the $r\times r$ nonsingular diagonal matrices of the singular values, and T_G and S_H are $r\times r$ matrices. Write

$$M:=\Sigma_G T_G\Sigma S_H\Sigma_H.$$

Then

$$M^{-1}=\Sigma_H^{-1}S_H^*\Sigma^{-1}T_G^*\Sigma_G^{-1},$$

and consequently

$$||M^{-1}||\le ||\Sigma_H^{-1}||\ ||S_H^*||\ ||\Sigma^{-1}||\ ||T_G^*||\ ||\Sigma_G^{-1}||.$$

Hence

$$||M^{-1}||\le ||\Sigma_H^{-1}||\ ||\Sigma^{-1}||\ |\Sigma_G^{-1}||$$

because S_H and T_G are unitary matrices. It follows from (12) for $W=M$ that

$$\sigma_r(M)\ge\sigma_r(G)\sigma_r(\Sigma)\sigma_r(H).$$

Now let $M=S_M\Sigma_M T_M$ be SVD where S_M and T_M are $r\times r$ unitary matrices.

Then $S:=S_GS_M$ and $T:=T_MT_H$ are unitary matrices, and so $G\Sigma H=S\Sigma_M T$ is SVD.

Therefore $\sigma_r(G\Sigma H)=\sigma_r(M)\ge\sigma_r(G)\sigma_r(\Sigma)\sigma_r(H)$. Combine this bound with (12) for W standing for G, Σ, H, and $G\Sigma H$. □

C The volume and r-projective volume of a perturbed matrix

Theorem 27. *Suppose that W' and E are $k \times l$ matrices,* $\text{rank}(W') = r \le \min\{k, l\}$, *$W = W' + E$, and $||E|| \le \epsilon$. Then*

$$\left(1 - \frac{\epsilon}{\sigma_r(W)}\right)^r \le \prod_{j=1}^{r}\left(1 - \frac{\epsilon}{\sigma_j(W)}\right) \le \frac{v_{2,r}(W)}{v_{2,r}(W')} \le \prod_{j=1}^{r}\left(1 + \frac{\epsilon}{\sigma_j(W)}\right) \le \left(1 + \frac{\epsilon}{\sigma_r(W)}\right)^r. \quad (13)$$

If $\min\{k, l\} = r$, *then $v_2(W) = v_{2,r}(W)$, $v_2(W') = v_{2,r}(W')$, and*

$$\left(1 - \frac{\epsilon}{\sigma_r(W)}\right)^r \le \frac{v_2(W)}{v_2(W')} = \frac{v_{2,r}(W)}{v_{2,r}(W')} \le \left(1 + \frac{\epsilon}{\sigma_r(W)}\right)^r. \quad (14)$$

Proof. Bounds (13) follow because a perturbation of a matrix within a norm bound ϵ changes its singular values by at most ϵ (see [GL13, Corollary 8.6.2]). Bounds (14) follow because $v_2(M) = v_{2,r}(M) = \prod_{j=1}^{r} \sigma_j(M)$ for any $k \times l$ matrix M with $\min\{k, l\} = r$, in particular for $M = W'$ and $M = W = W' + E$. □

If the ratio $\frac{\epsilon}{\sigma_r(W)}$ is small, then $\left(1 - \frac{\epsilon}{\sigma_r(W)}\right)^r = 1 - O\left(\frac{r\epsilon}{\sigma_r(W)}\right)$ and $\left(1 + \frac{\epsilon}{\sigma_r(W)}\right)^r = 1 + O\left(\frac{r\epsilon}{\sigma_r(W)}\right)$, which shows that the relative perturbation of the volume is amplified by at most a factor of r in comparison to the relative perturbation of the r largest singular values.

D The volume and r-projective volume of a matrix product

Theorem 28. [See Examples 29 and 30 below.]

Suppose that $W = GH$ for an $m \times q$ matrix G and a $q \times n$ matrix H. Then
(i) $v_2(W) = v_2(G)v_2(H)$ if $q = \min\{m, n\}$; $v_2(W) = 0 \le v_2(G)v_2(H)$ if $q < \min\{m, n\}$.
(ii) $v_{2,r}(W) \le v_{2,r}(G)v_{2,r}(H)$ for $1 \le r \le q$,
(iii) $v_2(W) \le v_2(G)v_2(H)$ if $m = n \le q$.

The following examples show some limitations on the extension of the theorem.

Example 29. If G and H are unitary matrices and if $GH = O$, then $v_2(G) = v_2(H) = v_{2,r}(G) = v_{2,r}(H) = 1$ and $v_2(GH) = v_{2,r}(GH) = 0$ for all $r \le q$.

Example 30. If $G = (1 \mid 0)$ and $H = \text{diag}(1, 0)$, then $v_2(G) = v_2(GH) = 1$ and $v_2(H) = 0$.

Proof. The theorem has been proved in [OZ18]. Next we include an alternative proof.

We first prove claim (i).

Let $G = S_G \Sigma_G T_G^*$ and $H = S_H \Sigma_H T_H^*$ be SVDs such that Σ_G, T_G^*, S_H, Σ_H, and $U = T_G^* S_H$ are $q \times q$ matrices and S_G, T_G^*, S_H, T_H^*, and U are unitary matrices.

Write $V := \Sigma_G U \Sigma_H$. Notice that $\det(V) = \det(\Sigma_G)\det(U)\det(\Sigma_H)$. Furthermore $|\det(U)| = 1$ because U is a square unitary matrix. Hence $v_2(V) = |\det(V)| = |\det(\Sigma_G)\det(\Sigma_H)| = v_2(G)v_2(H)$.

Now let $V = S_V \Sigma_V T_V^*$ be SVD where S_V, Σ_V, and T_V^* are $q \times q$ matrices and where S_V and T_V^* are unitary matrices.

Observe that $W = S_G V T_H^* = S_G S_V \Sigma_V T_V^* T_H^* = S_W \Sigma_V T_W^*$ where $S_W = S_G S_V$ and $T_W^* = T_V^* T_H^*$ are unitary matrices. Consequently $W = S_W \Sigma_V T_W^*$ is SVD, and so $\Sigma_W = \Sigma_V$.

Therefore $v_2(W) = v_2(V) = v_2(G)v_2(H)$ unless $q < \min\{m, n\}$. This proves claim (i) because clearly $v_2(W) = 0$ if $q < \min\{m, n\}$.

Next prove claim (ii).

First assume that $q \le \min\{m,n\}$ as in claim (i) and let $W = S_W \Sigma_W T_W^*$ be SVD.

In this case we have proven that $\Sigma_W = \Sigma_V$ for $V = \Sigma_G U \Sigma_H$, $q \times q$ diagonal matrices Σ_G and Σ_H, and a $q \times q$ unitary matrix U. Consequently $v_{2,r}(W) = v_{2,r}(\Sigma_V)$.

In order to prove claim (ii) in the case where $q \le \min\{m,n\}$, it remains to deduce that

$$v_{2,r}(\Sigma_V) \le v_{2,r}(G) v_{2,r}(H). \tag{15}$$

Notice that $\Sigma_V = S_V^* V T_V = S_V^* \Sigma_G U \Sigma_H T_V$ for $q \times q$ unitary matrices S_V^* and H_V.

Let $\Sigma_{r,V}$ denote the $r \times r$ leading submatrix of Σ_V, and so $\Sigma_{r,V} = \widehat{G}\widehat{H}$ where $\widehat{G} := S_{r,V}^* \Sigma_G U$ and $\widehat{H} := \Sigma_H T_{r,V}$ and where $S_{r,V}$ and $T_{r,V}$ denote the $r \times q$ leftmost unitary submatrices of the matrices S_V and T_V, respectively.

Observe that $\sigma_j(\widehat{G}) \le \sigma_j(G)$ for all j because $\widehat{G}$ is a submatrix of the $q \times q$ matrix $S_V^* \Sigma_G U$, and similarly $\sigma_j(\widehat{H}) \le \sigma_j(H)$ for all j. Therefore $v_{2,r}(\widehat{G}) = v_2(\widehat{G}) \le v_{2,r}(G)$ and $v_{2,r}(\widehat{H}) = v_2(\widehat{H}) \le v_{2,r}(H)$. Also notice that $v_{2,r}(\Sigma_{r,V}) = v_2(\Sigma_{r,V})$.

Furthermore $v_2(\Sigma_{r,V}) \le v_2(\widehat{G}) v_2(\widehat{H})$ by virtue of claim (i) because $\Sigma_{r,V} = \widehat{G}\widehat{H}$.

Combine the latter relationships and obtain (15), which implies claim (ii) in the case where $q \le \min\{m,n\}$.

Next we extend claim (ii) to the general case of any positive integer q.

Embed a matrix H into a $q \times q$ matrix $H' := (H \mid O)$ banded by zeros if $q > n$. Otherwise write $H' := H$. Likewise embed a matrix G into a $q \times q$ matrix $G' := (G^T \mid O)^T$ banded by zeros if $q > m$. Otherwise write $G' := G$.

Apply claim (ii) to the $m' \times q$ matrix G' and $q \times n'$ matrix H' where $q \le \min\{m', n'\}$.

Obtain that $v_{2,r}(G'H') \le v_{2,r}(G') v_{2,r}(H')$.

Substitute equations $v_{2,r}(G') = v_{2,r}(G)$, $v_{2,r}(H') = v_{2,r}(H)$, and $v_{2,r}(G'H') = v_{2,r}(GH)$, which hold because the embedding keeps invariant the singular values and therefore keeps invariant the volumes of the matrices G, H, and GH. This completes the proof of claim (ii), which implies claim (iii) because $v_2(V) = v_{2,n}(V)$ if V stands for G, H, or GH and if $m = n \le q$. $\square$

E Optimization of the sizes of CUR generators

Let us optimize the size $k \times l$ of a CUR generator towards minimization of the bounds of Theorems 4 and 5 on the error norm $||E||_C$.

The bound of Theorem 4 turns into

$$||E||_C \le (r+1)\ h\ \sigma_{r+1}(W)$$

if $k = l = r$ and into

$$||E||_C \le \sqrt{(1+1/b)(r+1)}\ h\ \sigma_{r+1}(W)$$

if $k = r = (b+1)l - 1$ or $l = r = (b+1)k - 1$ and if $b > 0$, that is, we decrease the output error bound by a factor of $\sqrt{\frac{r+1}{1+1/b}}$ in the latter case.

The bound of Theorem 5 turns into

$$||E||_C \le (1+1/b) h\ \sigma_{r+1}(W)$$

and is minimized for $k = l = (b+1)r - 1$ and a positive b.

Thus the volume is maximal where $\min\{k,l\} = r < \max\{k,l\}$, and the r-projective volume is maximal where $l = k > r$. The upper estimate of Theorem 5 for the norm $||E||_C$ converges to $\sigma_{r+1}(W)$ as $h \to 1$ and $b \to \infty$.

Acknowledgements: Our research was supported by NSF Grants CCF–1116736, CCF–1563942, and CCF–133834 and PSC CUNY Award 69813 00 48. We also thank A. Osinsky and N. L. Zamarashkin for pointers to their papers [OZ16] and [OZ18], S. A. Goreinov for reprints of his papers, and E. E. Tyrtyshnikov for pointers to the bibliography and the challenge of formally supporting empirical power of C–A algorithms.

References

[ABBB99] Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D., *LAPACK Users' Guide* (Third ed.), Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999. ISBN 0-89871-447-8

[B00] M. Bebendorf, Approximation of Boundary Element Matrices, *Numer. Math.*, **86, 4**, 565–589, 2000.

[B11] M. Bebendorf, Adaptive Cross Approximation of Multivariate Functions, *Constructive approximation*, **34, 2**, 149–179, 2011.

[BG65] P.A. Businger, G. H. Golub, Linear Least Squares Solutions by Householder Transformations, *Numer. Math.*, **7**, 269–276, 1965.

[B-I92] A. Ben-Israel, A Volume Associated with $m \times n$ Matrices, *Linear Algebra and its Applications*, **167**, 87–111, 1992.

[BKV15] M. Bebendorf, C. Kuske, R. Venn, Wideband Nested Cross Approximation for Helmholtz Problems, *Numerische Mathematik*, **130, 1**, 1–34, 2015.

[BR03] M. Bebendorf, S. Rjasanow, Adaptive Low-Rank Approximation of Collocation Matrices, *Computing*, **70, 1**, 1–24, 2003.

[BW17] C. Boutsidis, D. Woodruff, Optimal CUR Matrix Decompositions, *SIAM Journal on Computing*, **46, 2**, 543–589, 2017, DOI:10.1137/140977898.

[BW18] A. Bakshi, D. P. Woodruff: Sublinear Time Low-Rank Approximation of Distance Matrices, *Procs. 32nd Intern. Conf. Neural Information Processing Systems (NIPS'18),* 3786–3796, Montral, Canada, 2018.

[C87] T. F. Chan, Rank Revealing QR-Factorizations, *Lin. Algebra and Its Appls.*, **88/89**, 67–82, 1987.

[CBSW14] Y. Chen, S. Bhojanapalli, S. Sanghavi, R. Ward, Coherent Matrix Completion, *Proceedings of the 31st International Conference on Machine Learning*, PMLR **32, 1**, 674–682, 2014.

[CC70] J. D. Carroll, J. J. Chang, Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of Eckart-Young Decomposition, *Psychometrika*, **35**, 283–319, 1970.

[CH90] T.F. Chan, P.C. Hansen, Computing Truncated Singular Value Decomposition and Least Squares Solutions by Rank Revealing QR-factorizations, *SIAM Journal on Scientific and Statistical Computing*, **11**, 519 – 530, 1990.

[CH92] T.F. Chan, P.C. Hansen, Some Applications of the Rank Revealing QR Factorization, *SIAM Journal on Scientific and Statistical Computing*, **13**, 727 – 741, 1992.

[CI94] S. Chandrasekaran, I. Ipsen, On Rank Revealing QR Factorizations, *SIAM J. Matrix Anal. Appl.*, **15**, 592–622, 1994.

[DMM08] P. Drineas, M.W. Mahoney, S. Muthukrishnan, Relative-error CUR Matrix Decompositions, *SIAM Journal on Matrix Analysis and Applications*, **30, 2**, 844–881, 2008.

[G59] F. R. Gantmacher, *Theory of Matrices*, Chelsea, New York, 1959.

[G65] G. H. Golub, Numerical Methods for Solving Linear Least Squares Problem, *Numerische Math.*, **7**, 206–216, 1965.

[GE96] M. Gu, S.C. Eisenstat, An Efficient Algorithm for Computing a Strong Rank Revealing QR Factorization, *SIAM J. Sci. Comput.*, **17**, 848–869, 1996.

[GL13] G. H. Golub, C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Maryland, 2013 (fourth edition).

[GOSTZ10] S. Goreinov, I. Oseledets, D. Savostyanov, E. Tyrtyshnikov, N. Zamarashkin, How to Find a Good Submatrix, in *Matrix Methods: Theory, Algorithms, Applications* (dedicated to the Memory of Gene Golub, edited by V. Olshevsky and E. Tyrtyshnikov), pages 247–256, World Scientific Publishing, New Jersey, ISBN-13 978-981-283-601-4, ISBN-10-981-283-601-2, 2010.

[GT01] S. A. Goreinov, E. E. Tyrtyshnikov, The Maximal-Volume Concept in Approximation by Low Rank Matrices, *Contemporary Mathematics*, **208**, 47–51, 2001.

[GT11] S. A. Goreinov, E. E. Tyrtyshnikov, Quasioptimality of Skeleton Approximation of a Matrix on the Chebyshev Norm, *Russian Academy of Sciences: Doklady*, **83, 3**, 1–2, 2011.

[GTZ97] S. A. Goreinov, E. E. Tyrtyshnikov, N. L. Zamarashkin, A Theory of Pseudoskeleton Approximations, *Linear Algebra and Its Applications*, **261**, 1–21, 1997.

[GTZ97a] S. A. Goreinov, E. E. Tyrtyshnikov, N. L. Zamarashkin, Pseudo-skeleton Approximations by Matrices of Maximal Volume, *Mathematical Notes*, **62, 4**, 515–519, 1997.

[GZT95] S. A. Goreinov, N. L. Zamarashkin, E. E. Tyrtyshnikov, Pseudo-skeleton Approximations of Matrices, *Russian Academy of Sciences: Doklady*, **343, 2**, 151–152, 1995.

[H70] R. A. Harshman, Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multi-Modal Factor Analysis, *UCLA Working Papers in Phonetics,* **16**, 1–84, 1970.

[HLY92] T.-S. Hwang, W.-W. Lin, E.K. Yang, Rank Revealing LU Factorizations, *Linear Algebra and Its Applications*, **175**, 115–141, 1992.

[HMT11] N. Halko, P. G. Martinsson, J. A. Tropp, Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions, *SIAM Review*, **53, 2**, 217–288, 2011.

[HP92] Y.P. Hong, C.-T. Pan, Rank-Revealing QR Factorizations and the Singular Value Decomposition, *Math. of Computation*, **58**, 213–232, 1992.

[K85] D. E. Knuth, Semi-Optimal Bases for Linear Dependencies, *Linear and Multilinear Algebra*, **17, 1**, 1–4, 1985.

[KS16] N. Kishore Kumar, J. Schneider, Literature Survey on Low Rank Approximation of Matrices, arXiv:1606.06511v1 [math.NA] 21 June 2016.

[KV16] B. N. Khoromskij, A. Veit, Efficient Computation of Highly Oscillatory Itegrals by Using QTT Tensor Approximation, *Comp. Meth. Appl. Math.* **16, 1**, 145–159, 2016. Also available at arXiv:1408.5224, 2014.

[M11] M. W. Mahoney, Randomized Algorithms for Matrices and Data, *Foundations and Trends in Machine Learning*, NOW Publishers, **3, 2**, 2011. (Abridged version in: Advances in Machine Learning and Data Mining for Astronomy, edited by M. J. Way, et al., pp. 647–672, 2012.)

[M14] A. Y. Mikhalev, Method for the Low Rank Block Approximation of a Matrix from Its Entries, PhD Thesis, 105 pages, 2014.

[MD09] M. W. Mahoney, P. Drineas, CUR Matrix Decompositions for Improved Data Analysis, *Proc. Natl. Acad. Sci. USA*, **106**, 697–702, 2009.

[MW17] Cameron Musco, D. P. Woodruff: Sublinear Time Low-Rank Approximation of Positive Semidefinite Matrices, *IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS),* 672–683, 2017.

[OT10] I.V. Oseledets, E.E. Tyrtyshinkov, TT-Cross Approximation for Multi- Dimensional Arrays, *Linear Algebra and its Applications*, **432**, 70–88, 2010.

[OZ16] A.I. Osinsky, N. L. Zamarashkin, New Accuracy Estimates for Pseudo-skeleton Approximations of Matrices, Doklady Mathematics, **94, 3**, 643–645, 2016.

[OZ18] A.I. Osinsky, N. L. Zamarashkin, Pseudo-skeleton Approximations with Better Accuracy Estimates, *Linear Algebra and Its Applications*, **537**, 221–249, 2018.

[P00] C.-T. Pan, On the Existence and Computation of Rank-Revealing LU Factorizations, *Linear Algebra and its Applications*, **316**, 199–222, 2000.

[P18] V. Y. Pan, Some Recent and New Techniques for Superfast (Sublinear Cost) Low Rank Approximation, arXiv 1812.11406 (Submitted on 28 Dec, 2018).

[PLa] V. Y. Pan, Q. Luan, Refinement of Low Rank Approximation of a Matrix at Sublinear Cost, arXiv:1906.04223 (Submitted on 10 Jun 2019).

[PLSZ16] V. Y. Pan, Q. Luan, J. Svadlenka, L. Zhao, Primitive and Cynical Low Rank Approximation, Preprocessing and Extensions, arXiv 1611.01391v1 (Submitted on 3 Nov, 2016).

[PLSZ17] V. Y. Pan, Q. Luan, J. Svadlenka, L. Zhao, Superfast Accurate Approximation of Low Rank Matrices, arXiv 1710.07946v1 (Submitted on 22 Oct, 2017).

[PLSZa] V. Y. Pan, Q. Luan, J. Svadlenka, L. Zhao, CUR Low Rank Approximation at Sub-linear Cost, arXiv:1906.04112 (Submitted on 10 Jun 2019).

[PQY15] V. Y. Pan, G. Qian, X. Yan, Random Multipliers Numerically Stabilize Gaussian and Block Gaussian Elimination: Proofs and an Extension to Low-rank Approximation, *Linear Algebra and Its Applications*, **481**, 202–234, 2015.

[PZ17a] V. Y. Pan, L. Zhao, New Studies of Randomized Augmentation and Additive Preprocessing, *Linear Algebra and Its Applications*, **527**, 256–305, 2017.

[PZ17b] V. Y. Pan, L. Zhao, Numerically Safe Gaussian Elimination with No Pivoting, *Linear Algebra and Its Applications*, **527**, 349–383, 2017.

[SWZ17] Z. Song, D. P. Woodruff, P. Zhong, Low Rank Approximation with Entrywise l_1-Norm Error, *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC 2017)*, 688–701, ACM Press, New York, 2017.

[T96] E.E. Tyrtyshnikov, Mosaic-Skeleton Approximations, *Calcolo*, **33, 1**, 47–57, 1996.

[T00] E. Tyrtyshnikov, Incomplete Cross-Approximation in the Mosaic-Skeleton Method, *Computing*, **64**, 367–380, 2000.

[W14] D.P. Woodruff, Sketching As a Tool for Numerical Linear Algebra, *Foundations and Trends in Theoretical Computer Science*, **10, 1–2**, 1–157, 2014.