Sublinear Cost Low Rank Approximation via Subspace Sampling

Victor Y. Pan[1(✉)], Qi Luan[2], John Svadlenka[3], and Liang Zhao[1,3]

[1] Department of Computer Science,
Lehman College of the City University of New York, Bronx, NY 10468, USA
victor.pan@lehman.cuny.edu

[2] Program in Mathematics, The Graduate Center of the City University of New York, New York, NY 10036, USA
qi_luan@yahoo.com

[3] Program in Computer Science, The Graduate Center of the City University of New York, New York, NY 10036, USA
jsvadlenka@gradcenter.cuny.edu, lzhao1@gc.cuny.edu
http://comet.lehman.cuny.edu/vpan/

Abstract. Low Rank Approximation (LRA) of a matrix is a hot research subject, fundamental for Matrix and Tensor Computations and Big Data Mining and Analysis. Computations with LRA can be performed at *sublinear cost*, that is, by using much fewer memory cells and arithmetic operations than an input matrix has entries. Although every sublinear cost algorithm for LRA fails to approximate the worst case inputs, we prove that our sublinear cost variations of a popular subspace sampling algorithm output accurate LRA of a large class of inputs.

Namely, they do so with a high probability (*whp*) for a random input matrix that admits its LRA. In other papers we propose and analyze other sublinear cost algorithms for LRA and Linear Least Sqaures Regression. Our numerical tests are in good accordance with our formal results.

Keywords: Low-rank approximation · Sublinear cost · Subspace sampling

2000 Math. Subject Classification: 65Y20 · 65F30 · 68Q25 · 68W20 · 15A52

1 Introduction

LRA Background. Low rank approximation (LRA) of a matrix is a hot research area of Numerical Linear Algebra (NLA) and Computer Science (CS) with applications to fundamental matrix and tensor computations and Data Mining and Analysis (see surveys [HMT11,M11,KS17], and [CLO16]). Matrices from Big Data (e.g., unfolding matrices of multidimensional tensors) are frequently so immense that realistically one can access only a tiny fraction of

D. Slamanig et al. (Eds.): MACIS 2019, LNCS 11989, pp. 89–104, 2020.
https://doi.org/10.1007/978-3-030-43120-4_9

their entries, although quite typically these matrices admit their LRA (cf. (1) in Sect. 2). One can operate with such matrices *at sublinear computational cost*, that is, by using much fewer memory cells and arithmetic operations than an input matrix has entries, but can we compute LRA at sublinear cost? Yes and no. No, because every sublinear cost LRA algorithm fails even on the small input families of Appendix B. Yes, because our sublinear cost variations of a popular *subspace sampling* algorithm output accurate LRA for a large class of input.

Let us provide some details.

Subspace sampling algorithms compute LRA of a matrix M by using auxiliary matrices FM, MH or FMH for random multipliers F and H, commonly called *test matrices* and having smaller sizes. The output LRA are nearly optimal whp provided that F and H are Gaussian, Rademacher's, SRHT or SRFT matrices;[1] furthermore the algorithms consistently output accurate LRA in their worldwide application with these and some other random multipliers F and H, all of which, however, are multiplied by M at superlinear cost (see [TYUC17, Section 3.9], [HMT11, Section 7.4], and the bibliography therein).

Our modifications are deterministic. They use fixed sparse orthogonal (e.g., *subpermutation*) multipliers[2] F and H, run at sublinear cost, and whp output reasonably close *dual LRA*, i.e., LRA of a random input admitting LRA; we deduce our error estimates under three distinct models of random matrix computations in Sections 4.1 – 4.3. Unlike the customary randomized algorithms of [HMT11], [M11], [KS17], which perform at superlinear cost and which whp output close LRA of *any matrix* that admits LRA, our deterministic algorithms run at sublinear cost and whp output close LRA of *many such matrices* and in a sense most of them. Namely we prove that whp they output close LRA of a random input matrix that admits LRA.

How meaningful are our results? Our definitions of three classes of random matrices of low numerical rank are quite natural for various real world applications of LRA, but are odd for some other ones, as is the case with any definition of that kind. In spite of such odds, however, our formal study is in good accordance with our numerical tests for both synthetic and real world inputs, some from [HMT11]. Surely it is not realistic to assume that an input matrix is random, but we can randomize it by means of pre-processing of an input with random multipliers and then apply our results. Moreover, empirically such a randomized pre-processing and sublinear cost pre-processing with proper sparse multipliers consistently give similar results.

Our upper bounds on the output error of LRA of an $m \times n$ matrix of numerical rank r exceed the optimal error bound by a factor of $\sqrt{\min\{m,n\}r}$, but if the optimal bound is small enough we can apply two algorithms for iterative

[1] Here and hereafter *"Gaussian matrices"* stands for "Gaussian random matrices" (see Definition 1). "SRHT and SRFT" are the acronyms for "Subsample Random Hadamard and Fourier transforms". Rademacher's are the matrices filled with iid variables, each equal to 1 or -1 with probability $1/2$.

[2] Subpermutation matrices are full-rank submatrices of permutation matrices.

refinement of LRA, proposed in [PLa], running at sublinear cost, and reasonably efficient according to the results of numerical tests in [PLa].

As we discussed earlier, any sublinear cost LRA algorithm (and ours are no exception) fails on some families of hard inputs, but our analysis and tests show that the class of such inputs is narrow. We conjecture that it shrinks fast if we recursively apply the same algorithm with new multipliers; in Sect. 5 we comment of some heuristic recipes for these recursive processes; our numerical tests consistently confirm their efficiency.

Impact of Our Study, Its Extensions and By-Products

(i) Our duality approach is efficient for some fundamental matrix computations besides LRA: [PQY15, PZ17a], and [PZ17b] formally support empirical efficiency of dual Gaussian elimination with no pivoting, while [LPb] proposes a dual sublinear cost deterministic modification of Sarlós' randomized algorithm of 2006 and then proves that whp it outputs nearly optimal solution of the important problem of *Linear Least Squares Regression (LLSR)* for random input, and consequently for a large class of inputs – in a sense for most of them. This formal study turned out to be in very good accordance with the results of our extensive tests with synthetic and real world inputs.

(ii) In the paper [PLa] we proposed, analyzed, and tested new sublinear cost algorithms for refinement of a crude but reasonably close LRA.

(iii) In [LPa] and [PLSZa] we proved that popular Cross-Approximation LRA algorithms running at sublinear cost as well as our simplified sublinear cost variations of these algorithms output accurate solution of dual LRA whp, and we also devised a sublinear cost algorithm for transformation of any LRA into its special form of CUR LRA, which is particularly memory efficient.

(iv) Our acceleration of LRA can be immediately extended to the acceleration of Tensor Train Decomposition because it is reduced to recursive computation of LRA of unfolding matrices. Likewise our results can be readily extended to Tucker Decomposition of tensors because Tucker Decomposition is essentially LRA of unfolding matrices of a tensor. Extension to CP Decomposition of Tensors, however, remains a challenge.

(v) In [LPa] we also extended our progress by devising deterministic and practically promising algorithm that at sublinear cost computes accurate LRA for a symmetric positive semidefinite matrix admitting LRA.

Related Works. LRA has huge bibliography; see, e.g., [M11, HMT11, KS17]. The papers [PLSZ16] and [PLSZ17] have provided the first formal support for dual accurate randomized LRA at sublinear cost (they call sublinear cost algorithms *superfast*). The earlier papers [PQY15, PLSZ16, PZ17a], and [PZ17b] studied duality for other fundamental matrix computations besides LRA, and we have already cited extension of our progress in [PLa], [LPa] and [LPb].

Organization of the Paper. In Sect. 2 we recall random sampling for LRA. In Sects. 3 and 4 we estimate output errors of our dual LRA algorithms running at sublinear cost. In Sect. 5 we generate multipliers for both pre-processing and sampling. Appendix A is devoted to background on matrix computations. In Appendix B we specify some small families of inputs on which any sublinear cost LRA algorithm fails. Because of size limitation for this paper we leave to [PLSZb] various details, our historical comments, the test results, and some proofs, in particular the proofs of Theorems 5 and 6.

Some Definitions. The concepts "large", "small", "near", "close", "approximate", "ill-" and "well-conditioned", are usually quantified in the context. "≪" and "≫" mean "much less than" and "much greater than", respectively. *"Flop"* stands for "floating point arithmetic operation"; "*iid*" for "independent identically distributed". In context a *"perturbation of a matrix"* can mean a perturbation having a small relative norm. $\mathbb{R}^{p\times q}$ denotes the class of $p \times q$ real matrices. We assume dealing with real matrices throughout, and so the Hermitian transpose M^* of M turns into transpose M^T, but our study can be readily extended to complex matrices; see some relevant results about complex Gaussian matrices in [E88,CD05,ES05], and [TYUC17].

2 Four Known Subspace Sampling Algorithms

Hereafter $||\cdot||$ and $||\cdot||_F$ denote the spectral and the Frobenius matrix norms, respectively; $|\cdot|$ can denote either of them. M^+ denotes the Moore − Penrose pseudo inverse of M.

Next we devise a sublinear cost algorithm for LRA $XY$of matrix M such that

$$M = XY + E,\ ||E||/||M|| \le \epsilon, \tag{1}$$

for pairs of matrices X of size $m \times r$ and Y of size $r \times n$, a matrix norm $||\cdot||$, and a small tolerance ϵ.

Algorithm 1. *Range Finder* (see Remark 1).

INPUT: *An $m \times n$ matrix M and a target rank r.*
OUTPUT: *Two matrices $X \in \mathbb{R}^{m\times l}$ and $Y \in \mathbb{R}^{l\times m}$ defining an LRA $\tilde{M} = XY$..*
INITIALIZATION: *Fix an integer l, $r \le l \le n$, and an $n\times l$ test matrix (multiplier) H of rank l.*
COMPUTATIONS:

1. *Compute the $m \times l$ matrix MH.*
2. *Fix a nonsingular matrix $T^{-1} \in \mathbb{R}^{l\times l}$ and output the matrix $X := MHT^{-1} \in \mathbb{R}^{m\times l}$.*
3. *Output an $l \times n$ matrix $Y := \text{argmin}_V\ |XV - M| = X^+MT$.*

Remark 1. Let $\text{rank}(FM) = k$. Then $XY = MH(MH)^+M$ independently of the choice of T^{-1}, but a proper choice of a nonsingular matrix T numerically stabilizes the algorithm. For $l > r \ge \text{nrank}(MH)$ the matrix MH is ill-conditioned,[3] but let Q and R be the factors of the thin QR factorization

[3] nrank(W) denotes *numerical rank of W* (see Appendix A.1).

of MH, choose $T := R$, and observe that $X = MHT^{-1} = Q$ is an orthogonal matrix. $X = MHT^{-1}$ is also an orthogonal matrix if $T = R\Pi$ and if R and Π are factors of a rank-revealing QRΠ factorization of MH.

Column Subspace Sampling turns into *Column Subset Selection* in the case of a subpermutation matrix H.

Algorithm 2. *Transposed Range Finder* (see Remark 2).

INPUT: *As in Algorithm 1.*
OUTPUT: *Two matrices* $X \in \mathbb{R}^{k\times n}$ *and* $Y \in \mathbb{R}^{m\times k}$ *defining an LRA* $\tilde{M} = YX$.
INITIALIZATION: *Fix an integer* k, $r \le k \le m$, *and a* $k \times m$ *test matrix (multiplier)* F *of full numerical rank* k.
COMPUTATIONS:

1. *Compute the* $k \times m$ *matrix* FM.
2. *Fix a nonsingular* $k \times k$ *matrix* S^{-1}; *then output* $k \times n$ *matrix* $X := S^{-1}FM$.
3. *Output an* $m \times k$ *matrix* $Y := \operatorname{argmin}_V |VX - M|$.

Row Subspace Sampling turns into random *Row Subset Selection* in the case of a subpermutation matrix F.

Remark 2. $Y = M(S^{-1}FM)^+$ and $YX = M(FM)^+FM$ independently of the choice of S^{-1} if $\operatorname{rank}(FM) = l$, but a proper choice of S numerically stabilizes the algorithm. For $k > r \ge \operatorname{nrank}(FMH)$ the matrix FMH is ill-conditioned, but $S^{-1}FM$ is orthogonal if $S = L$, $X := Q = L^{-1}FM$, $Y := Q^*M$, and L and Q are the factors of the thin LQ factorization of FM.

The following algorithm combines row and column subspace sampling. In the case of the identity matrix S it turns into the algorithm of [TYUC17, Section 1.4], whose origin can be traced back to [WLRT08].

Algorithm 3. *Row and Column Subspace Sampling* (see Remark 3).

INPUT: *As in Algorithm 1.*
OUTPUT: *Two matrices* $X \in \mathbb{R}^{m\times k}$ *and* $Y \in \mathbb{R}^{k\times m}$ *defining an LRA* $\tilde{M} = XY$.
INITIALIZATION: *Fix two integers* k *and* l, $r \le k \le m$ *and* $r \le l \le n$; *fix two test matrices (multipliers)* $F \in \mathbb{R}^{k\times m}$ *and* $H \in \mathbb{R}^{n\times l}$ *of full numerical ranks and two nonsingular matrices* $S \in \mathbb{R}^{k\times k}$ *and* $T \in \mathbb{R}^{l\times l}$.
COMPUTATIONS:

1. *Output the matrix* $X = MHT^{-1} \in \mathbb{R}^{m\times l}$.
2. *Compute the matrices* $U := S^{-1}FM \in \mathbb{R}^{k\times n}$ *and* $W := S^{-1}FX \in \mathbb{R}^{m\times l}$.
3. *Output the* $l \times n$ *matrix* $Y := \operatorname{argmin}_V |W^+V - U|$.

Remark 3. $YX = MH(FMH)^+FM$ independently of the choice of the matrices S^{-1} and T^{-1} if the matrix FMH has full rank $\min\{k, l\}$, but a proper choice of S and T numerically stabilizes the computations of the algorithm. For $\min\{k, l\} > r \ge \operatorname{nrank}(FMH)$ the matrix FMH is ill-conditioned, but we can make it orthogonal by properly choosing the matrices S^{-1} and T^{-1}.

Remark 4. By applying Algorithm 3 to the transpose matrix M^* we obtain **Algorithm 4**, which begins with column subspace sampling followed by row subspace sampling. Our study of Algorithms 1 and 3 for input M actually covers Algorithms 2 and 4 as well.

Next we estimate the output errors of Algorithm 1 for any input; then extend these estimates to the output of Algorithm 3, at first for any input and then for random inputs.

3 Deterministic Error Bounds for Sampling Algorithms

Suppose that we are given matrices MHT^{-1} and $S^{-1}FM$. We can perform Algorithm 3 at arithmetic cost in $O(kln)$, which is sublinear if $kl \ll m$. Furthermore let $k^2 \ll m$ and $l^2 \ll n$. Then for proper deterministic choice of sparse (e.g., subpermutation) matrices S and T we can also compute the matrices MHT^{-1} and $S^{-1}FM$ at sublinear cost and thus complete computations of entire Algorithm 3 at sublinear cost. In this case we cannot ensure any reasonable accuracy of the output LRA for a worst case input and even for small input families of Appendix B, but we are going to prove that the output of that deterministic algorithm is quite accurate whp for random input and therefore for a large class of inputs, which is in good accordance with the results of our tests with synthetic and real world inputs.

We deduce some auxiliary deterministic output error bounds for any fixed input matrix in this section and refine them for random input under our probabilistic models in the next section. It turned out that the output error bounds are dominated at the stage of performing Range Finder because in Sect. 3.2 we rather readily bound additional impact of pre-processing with multipliers F and $S^{-1}F$.

3.1 Deterministic Error Bounds for Range Finder

Theorem 1 [HMT11, Theorem 9.1]**.** *Suppose that Algorithm 1 has been applied to a matrix M with a multiplier H and let*

$$C_1 = V_1^* H, \ C_2 = V_2^* H, \tag{2}$$

$$M = \begin{pmatrix} U_1 \ \Sigma_1 & & V_1^* \\ U_2 & \Sigma_2 & V_2^* \end{pmatrix}, \ M_r = U_1 \Sigma_1 V_1^*, \text{ and } M - M_r = U_2 \Sigma_2 V_2^* \tag{3}$$

be SVDs of the matrices M, its rank-r truncation M_r, and $M - M_r$, respectively. [$\Sigma_2 = O$ and $XY = M$ if $\text{rank}(M) = r$. The columns of V_1^* span the top right singular space of M.] Then

$$|M - XY|^2 \le |\Sigma_2|^2 + |\Sigma_2 C_2 C_1^+|^2. \tag{4}$$

Notice that $|\Sigma_2| = \bar{\sigma}_{r+1}(M)$, $|C_2| \le 1$, and $|\Sigma_2 C_2 C_1^+| \le |\Sigma_2| \ |C_2| \ |C_1^+|$ and obtain

$$|M - XY| \le (1 + |C_1^+|^2)^{1/2} \bar{\sigma}_{r+1}(M) \text{ for } C_1 = V_1^* H. \tag{5}$$

It follows that the output LRA is optimal up to a factor of $(1 + |C_1^+|^2)^{1/2}$.

Next we deduce an upper bound on the norm $|C_1^+|$ in terms of $||((MH)_r)^+||$, $||M||$, and $\eta := 2\sigma_{r+1}(M)\ ||((MH)_r)^+||$.

Corollary 1. *Under the assumptions of Theorem 1 let the matrix M_rH have full rank r. Then*

$$|(M_rH)^+|/|M_r^+| \leq |C_1^+| \leq |(M_rH)^+|\ |M_r| \leq |(M_rH)^+|\ |M|.$$

Proof. Deduce from (2) and (3) that $M_rH = U_1\Sigma_1C_1$. Hence $C_1 = \Sigma_1^{-1}U_1^*M_rH$.

Recall that the matrix M_rH has full rank r, apply Lemma 2, recall that U_1 is an orthogonal matrix, and obtain $|(M_rH)^+|/|\Sigma_1^{-1}| \leq |C_1^+| \leq |(M_rH)^+|\ |\Sigma_1|$.

Substitute $|\Sigma_1| = |M_r|$ and $|\Sigma_1^{-1}| = |M_r^+|$ and obtain the corollary.

Corollary 2. *See [PLSZb]. Under the assumptions of Corollary 1 let*

$$\eta := 2\sigma_{r+1}(M)\ ||((MH)_r)^+|| < 1,\ \eta' := \frac{2\sigma_{r+1}(M)}{1-\eta}\ ||((MH)_r)^+|| < 1.$$

Then

$$\frac{1-\eta'}{||M_r^+||}\ ||((MH)_r)^+|| \leq ||C_1^+\ || \leq \frac{||M||}{1-\eta}\ ||((MH)_r)^+||.$$

For a given matrix MH we compute the norm $||((MH)_r)^+||$ at sublinear cost if $l^2 \ll n$. If also some reasonable upper bounds on $||M||$ and $\sigma_{r+1}(M)$ are known, then Corollary 2 implies *a posteriori estimates* for the output errors of Algorithm 1.

3.2 Deterministic Impact of Pre-multiplication on the Errors of LRA

It turned out that the impact of pre-processing with multipliers $S^{-1}F$ into the output error bounds is dominated at the stage of Range Finder.

Lemma 1. [The impact of pre-multiplication on LRA errors.] *Suppose that Algorithm 3 outputs a matrix XY for $Y = (FX)^+FM$ and that $m \geq k \geq l = \text{rank}(X)$. Then*

$$M - XY = W(M - XX^+M) \text{ for } W = I_m - X(FX)^+F, \tag{6}$$

$$|M - XY| \leq |W|\ |M - XX^+M|,\ |W| \leq |I_m| + |X|\ |F|\ |(XF)^+|. \tag{7}$$

Proof. Recall that $Y = (FX)^+FM$ and notice that $(FX)^+FX = I_l$ if $k \geq l = \text{rank}(FX)$. Therefore $Y = X^+M + (FX)^+F(M - XX^+M)$. Consequently (6) and (7) hold.

We bounded the norm $|M - XX^+M|$ in the previous subsection; next we bound the norms $|(FX)^+|$ and $|W|$ of the matrices FX and W, computed at sublinear cost for $kl \ll n$, a fixed orthogonal X, and proper choice of sparse F.

Theorem 2. *[P00, Algorithm 1] for a real $h > 1$ applied to an $m \times l$ orthogonal matrix X performs $O(ml^2)$ flops and outputs an $l \times m$ subpermutation matrix F such that $||(FX)^+|| \le \sqrt{(m-l)lh^2+1}$, and $||W|| \le 1 + \sqrt{(m-l)lh^2+1}$, for $W = I_m + X(FX)^+F$ of (6) and any fixed $h > 1$; $||W|| \approx \sqrt{ml}$ for $m \gg l$ and $h \approx 1$.*

[P00, Algorithm 1] outputs $l \times m$ matrix F. One can strengthen deterministic bounds on the norm $|W|$ by computing proper $k \times m$ subpermutation matrices F for k of at least order l^2.

Theorem 3. *For k of at least order l^2 and a fixed orthogonal multiplier X compute a $k \times m$ subpermutation multiplier F by means of deterministic algorithms by Osinsky, running at sublinear cost and supporting [O18, equation (1)]. Then $||W|| \le 1 + ||(FX)^+|| = O(l)$ for W of (6).*

4 Accuracy of Sublinear Cost Dual LRA Algorithms

Next we estimate the output errors of Algorithm 1 for a fixed orthogonal matrix H and two classes of random inputs of low numerical rank, in particular for perturbed factor-Gaussian inputs of Definition 2. These estimates formally support the observed accuracy of Range Finder with various dense multipliers (see [HMT11, Section 7.4], and the bibliography therein), but also with sparse multipliers, with which Algorithms 3 and 4 run at sublinear cost.[4] We extend these upper estimates for output accuracy to variations of Algorithm 3 that run at sublinear cost; then we extend them to Algorithm 4 by means of transposition of an input matrix. This study involves the norms of a Gaussian matrix and its pseudo inverse, whose estimates we recall in Appendix A.4.

Hereafter $\stackrel{d}{=}$ denotes equality in probability distribution.

Definition 1. *A matrix is* Gaussian *if its entries are iid Gaussian (normal) variables. We let $\mathcal{G}^{p \times q}$ denote a $p \times q$ Gaussian matrices, and define random variables $\nu_{p,q} \stackrel{d}{=} |G|$, $\nu_{\mathrm{sp},p,q} \stackrel{d}{=} ||G||$, $\nu_{F,p,q} \stackrel{d}{=} ||G||_F$, $\nu^+_{p,q} \stackrel{d}{=} |G^+|$, $\nu^+_{\mathrm{sp},p,q} \stackrel{d}{=} ||G^+||$, and $\nu^+_{F,p,q} \stackrel{d}{=} ||G^+||_F$, for a $p \times q$ random Gaussian matrix G. [$\nu_{p,q} \stackrel{d}{=} \nu_{q,p}$ and $\nu^+_{p,q} \stackrel{d}{=} \nu^+_{q,p}$, for all pairs of p and q.]*

Theorem 4 [Non-degeneration of a Gaussian Matrix]**.** *Let $F \stackrel{d}{=} \mathcal{G}^{r \times p}$, $H \stackrel{d}{=} \mathcal{G}^{q \times r}$, $M \in \mathbb{R}^{p \times q}$ and $r \le \mathrm{rank}(M)$. Then the matrices F, H, FM, and MH have full rank r with probability 1.*

Assumption 1. *We simplify the statements of our results by assuming that a Gaussian matrix has full rank and ignoring the probability 0 of its degeneration.*

In Theorems 5 and 6 of the next subsections we state our error estimates, which we prove in [PLSZb].

[4] We defined Algorithm 4 in Remark 4

4.1 Errors of Range Finder for a Perturbed Factor-Gaussian Input

Assumption 2. *Suppose that $\tilde{M} = AB$ is a right $m \times n$ factor Gaussian matrix of rank r, $H = U_H \Sigma_H V_H^*$ is a $n \times l$ test matrix, and let $\theta = \frac{e\sqrt{l}(\sqrt{n}+\sqrt{r})}{l-r}$ be a constant. Here and hereafter $e := 2.71828182\ldots$. Define random variables $\nu = ||B||$ and $\mu = ||(BU_H)^+||$, and recall that $\nu \overset{d}{=} \nu_{sp,r,n}$ and $\mu \overset{d}{=} \nu_{sp,r,l}^+$.*

Theorem 5. [Errors of Range Finder for a perturbed factor-Gaussian matrix.] *Under Assumption 2, let $\phi = \left(\nu\mu||H^+||\right)^{-1} - 4\alpha||H||$, and let $M = \tilde{M} + E$ be a right factor Gaussian with perturbation such that*

$$\alpha := \frac{||E||_F}{(\sigma_r(M) - \sigma_{r+1}(M))} \leq \min\left(0.2, \frac{\xi}{8\kappa(H)\theta}\right) \tag{8}$$

where $0 < \xi < 2^{-0.5}$. Apply Algorithm 1 to M with a test matrix (multiplier) H. Then

$$||M - XY||^2 \leq \left(1 + \phi^{-2}\right)\sigma_{r+1}^2(M) \text{ and}$$

$$||M - XY|| \leq \left(1 + 2||H^+||\theta/\xi\right)\sigma_{r+1}(M) \tag{9}$$

with a probability no less than $1 - 2\sqrt{\xi}$. If $r \ll l$, then $\theta \approx e\sqrt{n/l}$, implying that the coefficient of $\sigma_{r+1}(M)$ on the right hand side of (9) is close to

$$1 + \frac{2e||H^+||}{\xi}\sqrt{n/l} = O(\sqrt{n/l}).$$

4.2 Output Errors of Range Finder Near a Matrix with a Random Singular Space

Next we state similar estimates under an alternative randomization model for dual LRA.

Theorem 6 [Errors of Range Finder for an input with a random singular space]**.** *Let the matrix V_1 in Theorem 1 be the $n \times r$ Q factor in a QR factorization of a normalized $n \times r$ Gaussian matrix G and let the multiplier $H = U_H \Sigma_H V_H^*$ be any $n \times l$ matrix of full rank $l \geq r$.*

(i) Then for random variables $\nu = |G|$ and $\mu = |G^T U_H|$, it holds that

$$|M - XY|/\bar{\sigma}_{r+1}(M) \leq \phi_{r,l,n} := (1 + (\nu\mu|H^+|)^2)^{1/2}.$$

(ii) For $n \geq l \geq r + 4 \geq 6$, with a probability at most $1 - 2\sqrt{\xi}$ it holds that

$$\phi_{sp,r,l,n}^2 \leq 1 + \xi^{-2}\ e^2\ ||H^+||^2 \Big(\frac{\sqrt{l}(\sqrt{n}+\sqrt{r})}{l-r}\Big)^2$$

and

$$\phi_{F,r,l,n}^2 \leq 1 + \xi^{-2}\ r^2\ ||H^+||_F^2\ \frac{n}{l-r-1}.$$

Here $||H^+|| = 1$ and $||H^+||_F = \sqrt{l}$ if the matrix H is orthogonal.

Bound the output errors of Algorithms 3 and 4 by combining the estimates of this section and Sect. 3.2 and by transposing an input matrix M.

4.3 Impact of Pre-multiplication in the Case of Gaussian Noise

Next deduce randomized estimates for the impact of pre-multiplication in the case where an input matrix M includes considerable additive white Gaussian noise,[5] which is a classical representation of natural noise in information theory, is widely adopted in signal and image processing, and in many cases properly represents the errors of measurement and rounding (cf. [SST06]).

Theorem 7. *Suppose that two matrices $F \in \mathbb{R}^{k\times m}$ and $H \in \mathbb{R}^{n\times l}$ are orthogonal where $k \geq 2l+2$, $l \geq 2$ and $k, l < \min(m,n)$, $A \in \mathbb{R}^{m\times n}$, λ_E is a positive scalar,*

$$M = A + E, \ \frac{1}{\lambda_E}\ E \stackrel{d}{=} \mathcal{G}^{m\times n}, \tag{10}$$

and $W = I_m - MH(FMH)^+F$ (cf. (6) for $X = MH$). Then

$$\mathbb{E}\Big(\frac{||W||_F - \sqrt{m}}{\lambda_E\ ||M||_F}\Big) \leq \sqrt{\frac{l}{k-2l-1}} \ \ \text{and} \ \ \mathbb{E}\Big(\frac{||W||-1}{\lambda_E\ ||M||}\Big) \leq \frac{e\sqrt{k-l}}{k-2l}. \tag{11}$$

Proof. Assumption (10) and Lemma 5 together imply that FEH is a scaled Gaussian matrix: $\frac{1}{\lambda_E}FEH \stackrel{d}{=} \mathcal{G}^{k\times l}$. Hence $FMH = FAH + \lambda_E G_{k,l}$. Apply Theorem 10 and obtain

$$\mathbb{E}\ ||(FMH)^+|| \leq \lambda_E \frac{e\sqrt{k-l}}{k-2l} \ \ \text{and} \ \ \mathbb{E}\ ||(FMH)^+||_F \leq \lambda_E \sqrt{\frac{l}{k-2l-1}}$$

Recall from (6) that $|W| \leq |I_m| + |(FMH)^+|\ |M|$ since the multipliers F and H are orthogonal, and thus

$$\mathbb{E}|W| \leq |I_m| + |M| \cdot \mathbb{E}\ |(FMH)^+|.$$

Substitute equations $||I_m||_F = \sqrt{m}$ and $||I_m|| = 1$ and claim (iii) of Theorem 12 and obtain (11).

Remark 5. For $k = l = \rho$, $S = T = I_k$, subpermutation matrices F and H, and a nonsingular matrix FMH, Algorithms 3 and 4 output LRA in the form CUR where $C \in \mathbb{R}^{m\times\rho}$ and $R \in \mathbb{R}^{\rho\times n}$ are two submatrices made up of ρ columns and ρ rows of M and $U = (FMH)^{-1}$. [PLSZa] extends our current study to devising and analyzing algorithms for the computation of such CUR LRA in the case where k and l are arbitrary integers not exceeded by ρ.

[5] Additive white Gaussian noise is statistical noise having a probability density function (PDF) equal to that of the Gaussian (normal) distribution.

5 Multiplicative Pre-processing for LRA

We proved that sublinear cost variations of Algorithms 3 and 4 whp output accurate LRA of a random input. In the real world computations input matrices are not random, but we can randomize them by multiplying them by random matrices.

Algorithms 1–4 output accurate LRA whp if such multipliers are Gaussian, SRHT, SRFT or Rademacher's (cf. [HMT11, Sections 10 and 11], [T11]. Multiplication by these matrices runs at a superlinear cost, and our heuristic recipe is to apply these algorithms with a small variety of sparse multipliers F_i and/or H_i, $i = 1, 2, \ldots$, with which computational cost becomes sublinear, and then to monitor the accuracy of the output LRA by applying the criteria of the previous section, [PLa], and/or [PLSZa].

Various families of sparse multipliers have been proposed, extensively tested in [PLSZ16] and [PLSZ17], and turned out to be nearly as efficient as Gaussian multpliers according to these tests. One can readily complement these families with subpermutation matrices and, say, sparse quasi Rademacher's multipliers (see [PLSZa]) and then combine these basic multipliers together into their orthogonalized sums, products or other lower degree polynomials (cf. [HMT11, Remark 4.6]).

Acknowledgements. We were supported by NSF Grants CCF–1116736, CCF–1563942, CCF–1733834 and PSC CUNY Award 69813 00 48.

Appendix

A Background on Matrix Computations

A.1 Some Definitions

- An $m \times n$ matrix M is *orthogonal* if $M^*M = I_n$ or $MM^* = I_m$.
- For $M = (m_{i,j})_{i,j=1}^{m,n}$ and two sets $\mathcal{I} \subseteq \{1, \ldots, m\}$ and $\mathcal{J} \subseteq \{1, \ldots, n\}$, define the submatrices $M_{\mathcal{I},:} := (m_{i,j})_{i \in \mathcal{I}; j=1,\ldots,n}$, $M_{:,\mathcal{J}} := (m_{i,j})_{i=1,\ldots,m; j \in \mathcal{J}}$, and $M_{\mathcal{I},\mathcal{J}} := (m_{i,j})_{i \in \mathcal{I}; j \in \mathcal{J}}$.
- $\text{rank}(M)$ denotes the *rank* of a matrix M.
- $\text{argmin}_{|E| \le \epsilon |M|} \text{rank}(M+E)$ is the ϵ-rank(M) it is *numerical rank*, nrank(M), if ϵ is small in context.
- Write $\sigma_j(M) = 0$ for $j > r$ and obtain M_r, the *rank-r truncation* of M.
- $\kappa(M) = ||M||\ ||M^+||$ is the spectral *condition number* of M.

A.2 Auxiliary Results

Next we recall some relevant auxiliary results (we omit the proofs of two well-known lemmas).

Lemma 2 [The norm of the pseudo inverse of a matrix product]. *Suppose that* $A \in \mathbb{R}^{k\times r}$, $B \in \mathbb{R}^{r\times l}$ *and the matrices* A *and* B *have full rank* $r \le \min\{k, l\}$. *Then* $|(AB)^+| \le |A^+|\ |B^+|$.

Lemma 3 (The norm of the pseudo inverse of a perturbed matrix, [B15, Theorem 2.2.4]). *If* $\text{rank}(M + E) = \text{rank}(M) = r$ *and* $\eta = ||M^+||\ ||E|| < 1$, *then*

$$\frac{1}{\sqrt{r}}||(M+E)^+|| \le ||(M+E)^+|| \le \frac{1}{1-\eta}||M^+||.$$

Lemma 4 (The impact of a perturbation of a matrix on its singular values, [GL13, Corollary 8.6.2]). *For* $m \ge n$ *and a pair of* $m \times n$ *matrices* M *and* $M+E$ *it holds that*

$$|\sigma_j(M+E) - \sigma_j(M)| \le ||E|| \text{ for } j = 1, \dots, n.$$

Theorem 8 (The impact of a perturbation of a matrix on its top singular spaces, [GL13, Theorem 8.6.5]). *Let* $g =: \sigma_r(M) - \sigma_{r+1}(M) > 0$ and $||E||_F \le 0.2g$. *Then for the left and right singular spaces associated with the* r *largest singular values of the matrices* M *and* $M + E$, *there exist orthogonal matrix bases* $B_{r,\text{left}}(M)$, $B_{r,\text{right}}(M)$, $B_{r,\text{left}}(M + E)$, *and* $B_{r,\text{right}}(M + E)$ *such that*

$$\max\{||B_{r,\text{left}}(M+E)-B_{r,\text{left}}(M)||_F, ||B_{r,\text{right}}(M+E)-B_{r,\text{right}}(M)||_F\} \le \frac{4||E||_F}{g}.$$

For example, if $\sigma_r(M) \ge 2\sigma_{r+1}(M)$, which implies that $g \ge 0.5\ \sigma_r(M)$, and if $||E||_F \le 0.1\ \sigma_r(M)$, then the upper bound on the right-hand side is approximately $8||E||_F/\sigma_r(M)$.

A.3 Gaussian and Factor-Gaussian Matrices of Low Rank and Low Numerical Rank

Lemma 5 [Orthogonal invariance of a Gaussian matrix]. *Suppose that* k, m, *and* n *are three positive integers,* $k \le \min\{m, n\}$, $G_{m,n} \overset{d}{=} \mathcal{G}^{m\times n}$, $S \in \mathbb{R}^{k\times m}$, $T \in \mathbb{R}^{n\times k}$, *and* S *and* T *are orthogonal matrices. Then* SG *and* GT *are Gaussian matrices.*

Definition 2 [Factor-Gaussian matrices]. *Let* $r \le \min\{m, n\}$ *and let* $\mathcal{G}^{m\times n}_{r,B}$, $\mathcal{G}^{m\times n}_{A,r}$, *and* $\mathcal{G}^{m\times n}_{r,C}$ *denote the classes of matrices* $G_{m,r}B$, $AG_{r,n}$, *and* $G_{m,r}CG_{r,n}$, *respectively, which we call* left, right, *and* two-sided factor-Gaussian matrices of rank r, *respectively, provided that* $G_{p,q}$ *denotes a* $p \times q$ *Gaussian matrix,* $A \in \mathbb{R}^{m\times r}$, $B \in \mathbb{R}^{r\times n}$, *and* $C \in \mathbb{R}^{r\times r}$, *and* A, B *and* C *are well-conditioned matrices of full rank* r.

Theorem 9. *The class* $\mathcal{G}^{m\times n}_{r,C}$ *of two-sided* $m \times n$ *factor-Gaussian matrices* $G_{m,r}\Sigma G_{r,n}$ *does not change if in its definition we replace the factor* C *by a well-conditioned diagonal matrix* $\Sigma = (\sigma_j)_{j=1}^r$ *such that* $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$.

Proof. Let $C = U_C \Sigma_C V_C^*$ be SVD. Then $A = G_{m,r} U_C \overset{d}{=} \mathcal{G}^{m\times r}$ and $B = V_C^* G_{r,n} \overset{d}{=} \mathcal{G}^{r\times n}$ by virtue of Lemma 5, and so $G_{m,r} C G_{r,n} = A \Sigma_C B$ for $A \overset{d}{=} \mathcal{G}^{m\times r}$, $B \overset{d}{=} \mathcal{G}^{r\times n}$, and A independent from B.

Definition 3. The relative norm of a perturbation of a Gaussian matrix *is the ratio of the perturbation norm and the expected value of the norm of the matrix (estimated in Theorem 11).*

We refer to all three matrix classes above as *factor-Gaussian matrices of rank* r, to their perturbations within a relative norm bound ϵ as *factor-Gaussian matrices of* ϵ*-rank* r, and to their perturbations within a small relative norm as *factor-Gaussian matrices of numerical rank* r to which we also refer as *perturbations of factor-Gaussian matrices.*

Clearly $||(A\Sigma)^+|| \le ||\Sigma^{-1}||\ ||A^+||$ and $||(\Sigma B)^+|| \le ||\Sigma^{-1}||\ ||B^+||$ for a two-sided factor-Gaussian matrix $M = A\Sigma B$ of rank r of Definition 2, and so whp such a matrix is both left and right factor-Gaussian of rank r.

Theorem 10. *Suppose that* λ *is a positive scalar,* $M_{k,l} \in \mathbb{R}^{k\times l}$ *and* G *a* $k \times l$ *Gaussian matrix for* $k - l \ge l + 2 \ge 4$*. Then, we have*

$$\mathbb{E}\ ||(M_{k,l} + \lambda G)^+|| \le \frac{\lambda e\sqrt{k-l}}{k-2l} \quad \textit{and} \quad \mathbb{E}\ ||(M_{k,l} + \lambda G)^+||_F \le \lambda\sqrt{\frac{l}{k-2l-1}}$$

Proof. Let $M_{k,l} = U\Sigma V^*$ be *full SVD* such that $U \in \mathbb{R}^{k\times k}$, $V \in \mathbb{R}^{l\times l}$, U and V are orthogonal matrices, $\Sigma = (D \mid O_{l,k-l})^*$, and D is an $l \times l$ diagonal matrix. Write $W_{k,l} := U^*(M_{k,l} + \lambda G)V$ and observe that $U^* M_{k,l} V = \Sigma$ and $U^*GV = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$ is a $k \times l$ Gaussian matrix by virtue of Lemma 5. Hence

$$\sigma_l(W_{k,l}) = \sigma_l\Big(\begin{bmatrix} D + \lambda G_1 \\ \lambda G_2 \end{bmatrix}\Big) \ge \max\{\sigma_l(D + \lambda G_1), \lambda\sigma_l(G_2)\},$$

and so $|W_{k,l}^+| \le \min\{|(D + \lambda G_1)^+|, |\lambda G_2^+|\}$. Recall that $G_1 \overset{d}{=} \mathcal{G}^{l\times l}$ and $G_2 \overset{d}{=} \mathcal{G}^{k-l\times l}$ are independent, and now Theorem 10 follows because $|(M_{k,l}+\lambda G_{k,l})^+| = |W_{k,l}^+|$ and by virtue of claim (iii) and (iv) of Theorem 12.

A.4 Norms of a Gaussian Matrix and Its Pseudo Inverse

$\Gamma(x) = \int_0^\infty \exp(-t)t^{x-1}dt$ denotes the Gamma function.

Theorem 11 [Norms of a Gaussian matrix. See [DS01, Theorem II.7] and our Definition 1].

(i) Probability$\{\nu_{\mathrm{sp},m,n} > t+\sqrt{m}+\sqrt{n}\} \le \exp(-t^2/2)$ *for* $t \ge 0$, $\mathbb{E}(\nu_{\mathrm{sp},m,n}) \le \sqrt{m} + \sqrt{n}$.

(ii) $\nu_{F,m,n}$ *is the* χ*-function, with* $\mathbb{E}(\nu_{F,m,n}) = mn$ *and probability density* $\frac{2x^{n-i}\exp(-x^2/2)}{2^{n/2}\Gamma(n/2)}$.

Theorem 12 [Norms of the pseudo inverse of a Gaussian matrix (see Definition 1)].

(i) Probability $\{\nu^+_{\mathrm{sp},m,n} \geq m/x^2\} < \frac{x^{m-n+1}}{\Gamma(m-n+2)}$ *for* $m \geq n \geq 2$ *and all positive* x,

(ii) Probability $\{\nu^+_{F,m,n} \geq t\sqrt{\frac{3n}{m-n+1}}\} \leq t^{n-m}$ *and* Probability $\{\nu^+_{\mathrm{sp},m,n} \geq t\frac{e\sqrt{m}}{m-n+1}\} \leq t^{n-m}$ *for all* $t \geq 1$ *provided that* $m \geq 4$,

(iii) $\mathbb{E}((\nu^+_{F,m,n})^2) = \frac{n}{m-n-1}$ *and* $\mathbb{E}(\nu^+_{\mathrm{sp},m,n}) \leq \frac{e\sqrt{m}}{m-n}$ *provided that* $m \geq n+2 \geq 4$,

(iv) Probability $\{\nu^+_{\mathrm{sp},n,n} \geq x\} \leq \frac{2.35\sqrt{n}}{x}$ *for* $n \geq 2$ *and all positive* x, *and furthermore* $||M_{n,n} + G_{n,n}||^+ \leq \nu_{n,n}$ *for any* $n \times n$ *matrix* $M_{n,n}$ *and an* $n \times n$ *Gaussian matrix* $G_{n,n}$.

Proof. See [CD05, Proof of Lemma 4.1] for claim (i), [HMT11, Proposition 10.4 and equations (10.3) and (10.4)] for claims (ii) and (iii), and [SST06, Theorem 3.3] for claim (iv).

Theorem 12 implies reasonable probabilistic upper bounds on the norm $\nu^+_{m,n}$ even where the integer $|m-n|$ is close to 0; whp the upper bounds of Theorem 12 on the norm $\nu^+_{m,n}$ decrease very fast as the difference $|m-n|$ grows from 1.

B Small Families of Hard Inputs for Sublinear Cost LRA

Any sublinear cost LRA algorithm fails on the following small families of LRA inputs.

Example 1. Let $\Delta_{i,j}$ denote an $m \times n$ matrix of rank 1 filled with 0s except for its (i,j)th entry filled with 1. The mn such matrices $\{\Delta_{i,j}\}_{i,j=1}^{m,n}$ form a family of δ*-matrices*. We also include the $m \times n$ null matrix $O_{m,n}$ filled with 0s into this family. Now any fixed sublinear cost algorithm 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 $1/2$. Arrive at the same conclusion by applying 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 sublinear cost.

The example actually covers randomized LRA algorithms as well. Indeed suppose that with a positive constant probability an LRA algorithm does not access K entries of an input matrix with a positive constant probability. Apply this algorithm to two matrices of low rank whose difference at all these K entries is equal to a large constant C. Then, clearly, with a positive constant probability the algorithm has errors at least $C/2$ at at least $K/2$ of these entries. The paper [LPa] shows, however, that accurate LRA of a matrix that admits sufficiently close

LRA can be computed at sublinear cost in two successive Cross-Approximation (C-A) iterations (cf. [GOSTZ10]) provided that we avoid choosing degenerating initial submatrix, which is precisely the problem with the matrix families of Example 1. Thus we readily compute close LRA if we recursively perform C-A iterations and avoid degeneracy at some C-A step.

References

[B15] Björck, Å.: Numerical Methods in Matrix Computations. TAM, vol. 59. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-05089-8

[CD05] Chen, Z., Dongarra, J.J.: Condition numbers of Gaussian random matrices, SIAM. J. Matrix Anal. Appl. **27**, 603–620 (2005)

[CLO16] Cichocki, C., Lee, N., Oseledets, I., Phan, A.-H., Zhao, Q., Mandic, D.P.: Tensor networks for dimensionality reduction and large-scale optimization: part 1 low-rank tensor decompositions. Found. Trends® Mach. Learn. **9**(4–5), 249–429 (2016)

[DS01] Davidson, K.R., Szarek, S.J.: Local operator theory, random matrices, and banach spaces. In: Johnson, W.B., Lindenstrauss, J., (eds.) Handbook on Geometry of Banach Spaces, pp. 317–368, North Holland (2001)

[E88] Edelman, A.: Eigenvalues and condition numbers of random matrices. SIAM J. Matrix Anal. Appl. **9**(4), 543–560 (1988)

[ES05] Edelman, A., Sutton, B.D.: Tails of condition number distributions. SIAM J. Matrix Anal. Appl. **27**(2), 547–560 (2005)

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

[GOSTZ10] Goreinov, S., Oseledets, I., Savostyanov, D., Tyrtyshnikov, E., Zamarashkin, N.: 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), pp. 247–256. World Scientific Publishing, New Jersey (2010)

[HMT11] Halko, N., Martinsson, P.G., Tropp, J.A.: Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. **53**(2), 217–288 (2011)

[KS17] Kishore Kumar, N., Schneider, J.: Literature survey on low rank approximation of matrices. Linear Multilinear Algebra **65**(11), 2212–2244 (2017). arXiv:1606.06511v1 [math.NA] 21 June 2016

[LPa] Luan, Q., Pan, V.Y.: CUR LRA at sublinear cost based on volume maximization, In: Salmanig, D. et al. (eds.) MACIS 2019, LNCS 11989, pp. xx–yy. Springer, Switzerland (2020). https://doi.org/10.1007/978-3-030-43120-49. arXiv:1907.10481 (2019)

[LPb] Luan, Q., Pan, V.Y., Randomized approximation of linear least squares regression at sublinear cost. arXiv:1906.03784, 10 June 2019

[M11] Mahoney, M.W.: Randomized algorithms for matrices and data. Found. Trends Mach. Learn. **3**, 2 (2011)

[O18] Osinsky, A.: Rectangular maximum volume and projective volume search algorithms. arXiv:1809.02334, September 2018

[P00] Pan, C.-T.: On the existence and computation of rank-revealing LU factorizations. Linear Algebra Appl. **316**, 199–222 (2000)

[PLa] Pan, V.Y., Luan, Q.: Refinement of low rank approximation of a matrix at sublinear cost. arXiv:1906.04223, 10 June 2019

[PLSZ16] Pan, V.Y., Luan, Q., Svadlenka, J., Zhao, L.: Primitive and Cynical Low Rank Approximation, Preprocessing and Extensions. arXiv 1611.01391, 3 November 2016

[PLSZ17] Pan, V.Y., Luan, Q., Svadlenka, J., Zhao, L.: Superfast Accurate Low Rank Approximation. Preprint, arXiv:1710.07946, 22 October 2017

[PLSZa] Pan, V.Y., Luan, Svadlenka, Q., Zhao, L.: CUR Low Rank Approximation at Sublinear Cost. arXiv:1906.04112, 10 June 2019

[PLSZb] Pan, V.Y., Luan, Q., Svadlenka, J., Zhao, L.: Low rank approximation at sublinear cost by means of subspace sampling. arXiv:1906.04327, 10 June 2019

[PQY15] Pan, V.Y., Qian, G., Yan, X.: Random multipliers numerically stabilize Gaussian and block Gaussian elimination: proofs and an extension to low-rank approximation. Linear Algebra Appl. **481**, 202–234 (2015)

[PZ17a] Pan, V.Y., Zhao, L.: New studies of randomized augmentation and additive preprocessing. Linear Algebra Appl. **527**, 256–305 (2017)

[PZ17b] Pan, V.Y., Zhao, L.: Numerically safe Gaussian elimination with no pivoting. Linear Algebra Appl. **527**, 349–383 (2017)

[SST06] Sankar, A., Spielman, D., Teng, S.-H.: Smoothed analysis of the condition numbers and growth factors of matrices. SIMAX **28**(2), 446–476 (2006)

[T11] Tropp, J.A.: Improved analysis of subsampled randomized Hadamard transform. Adv. Adapt. Data Anal. **3**(1–2), 115–126 (2011). (Special issue "Sparse Representation of Data and Images")

[TYUC17] Tropp, J.A., Yurtsever, A., Udell, M., Cevher, V.: Practical sketching algorithms for low-rank matrix approximation. SIAM J. Matrix Anal. Appl. **38**, 1454–1485 (2017)

[WLRT08] Woolfe, F., Liberty, E., Rokhlin, V., Tygert, M.: A fast randomized algorithm for the approximation of matrices. Appl. Comput. Harmonic. Anal. **25**, 335–366 (2008)