2012

TR-2012014: More on the Power of Randomized Matrix Multiplication

Victor Y. Pan

Guoliang Qian

More on the Power of Randomized Matrix Multiplication *

Victor Y. Pan[1,2],[a] and Guoliang Qian[2],[b]

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

Abstract

A random matrix is likely to be well conditioned, and motivated by this well known property we employ random matrix multipliers to advance some fundamental matrix computations. This includes numerical stabilization of Gaussian elimination with no pivoting as well as block Gaussian elimination, approximation of the leading and trailing singular spaces of an ill conditioned matrix, associated with its largest and smallest singular values, respectively, and approximation of this matrix by low-rank matrices, with further extensions to the approximation of tensor decomposition. We formally support the efficiency of the proposed techniques where we employ Gaussian random multipliers, but our extensive tests have consistently produced the same outcome where instead we used sparse and structured random multipliers, defined by much fewer random parameters compared to the number of their entries.

2000 Math. Subject Classification: 15A52, 15A12, 15A06, 65F22, 65F05

Key Words: Random matrices, GENP, Low-rank approximation, Numerical rank

1 Introduction

It is well known that A random matrix is likely to be well conditioned [D88], [E88], [ES05], [CD05], [SST06], [B11], and motivated by this well known property we apply randomized matrix multiplication to advance some fundamental matrix computations. We stabilize numerically Gaussian elimination with no pivoting as well as block Gaussian elimination, approximate leading and trailing singular spaces of an ill conditioned matrix A, associated with its largest and smallest singular values, respectively, approximate this matrix by low-rank matrices, and compute a Tensor Train approximation of a tensor as well as approximation of a matrix by a structured matrix lying nearby.

*Some results of this paper have been presented at the ACM-SIGSAM International Symposium on Symbolic and Algebraic Computation (ISSAC '2011), San Jose, CA, 2011, the 3nd International Conference on Matrix Methods in Mathematics and Applications (MMMA 2011) in Moscow, Russia, June 22-25, 2011, the 7th International Congress on Industrial and Applied Mathematics (ICIAM 2011), in Vancouver, British Columbia, Canada, July 18-22, 2011, the SIAM International Conference on Linear Algebra, in Valencia, Spain, June 18-22, 2012, and the Conference on Structured Linear and Multilinear Algebra Problems (SLA2012), in Leuven, Belgium, September 10-14, 2012

Our numerical tests are in good accordance with our formal study, except that in our tests all algorithms have fully preserved their power even where we dramatically decreased the number of random parameters involved by using sparse and structured multipliers.

1.1 Numerically safe Gaussian elimination with no pivoting

Hereafter "flop" stands for "arithmetic operation", by saying "expect" and "likely" we mean "with probability 1 or close to 1", $\sigma_j(A)$ denotes the jth largest singular value of an $n \times n$ matrix A, and the ratio $\kappa(A) = \sigma_1(A)/\sigma_\rho(A)$ for $\rho = \text{rank}(A)$ denotes its condition number. $\kappa(A) = ||A||\ ||A^{-1}||$ if $\rho = n$, that is if A is a nonsingular matrix. If this number is large in context, then the matrix A is *ill conditioned*, otherwise *well conditioned*. For matrix inversion and solving linear systems of equations the condition number represents the output magnification of input errors,

$$\kappa(A) \approx \frac{||\text{OUTPUT ERROR}||}{||\text{INPUT ERROR}||}, \tag{1.1}$$

and backward error analysis implies similar magnification of rounding errors [GL96], [H02], [S98].

To avoid dealing with singular or ill conditioned matrices in Gaussian elimination, one incorporates pivoting, that is row or column interchange. *Gaussian elimination with no pivoting* (hereafter we refer to it as *GENP*) can easily fail in numerical computations with rounding errors, except for the cases where the input matrices are strongly well conditioned, that is where all their leading principal square blocks are nonsingular and well conditioned. In particular diagonally dominant as well as positive definite well conditioned matrices have this property. For such matrices, GENP outperforms Gaussian elimination with pivoting [GL96, page 119]. Random matrices are likely to be strongly well conditioned, but we do not solve random linear systems of equations. We can, however randomize linear systems by applying random multipliers and then can apply GENP. We proposed and tested this approach in [PGMQ, Section 12.2] and [PQZa], and our tests consistently showed its efficiency even where we used just circulant or Householder multipliers filled with integers ± 1 and where we limited randomization to the choice of the signs $\pm$ (see our Table 7.1 and [PQZa, Table 2]). Our Corollary 4.1 supports these empirical observations provided that the multipliers are square Gaussian random matrices. Formal study of structured preconditioning was stated as a challenge in [SST06], and the problem remains largely open (see some recent advance in [PQa]).

1.2 Randomized low-rank approximation and beyond

Our Corollary 4.1, supporting randomized GENP, relies on the probabilistic estimates for the ranks and condition numbers of the products $\kappa(GA)$ and $\kappa(AH)$ in terms of $\kappa(A)$ where G and H are Gaussian random matrices (see Theorem 4.1). We also apply the same estimates to support randomized algorithms for the approximation of the leading singular spaces of an ill conditioned matrix A associated with its largest singular values. This can be immediately extended to the approximation of a matrix having a small numerical rank by low-rank matrices. The algorithm is numerially safe, runs at a low computational cost, and has a great number of highly important applications to matrix computations [HMT11]. We point out its further extensions to the approximation of a matrix by a structured matrix lying nearby and to computing a Tensor Train approximation of a tensor. Then again our formal support of these algorithms relies on using Gaussian random multipliers, but our tests show that random Toeplitz multipliers are as effective. This suggests formal and experimantal study of various other random structured and sparse multipliers that depend on smaller numbers of random parameters. Note the recent success of Tropp [T11] in this direction.

1.3 Related work

Preconditioning of linear systems of equations is a classical subject [A94], [B02], [G97]. Randomized multiplicative preconditioning for numerical stabilization of GENP was proposed in [PGMQ, Section 12.2] and [PQZa], but with no formal support for this approach. On low-rank approximation we refer the reader to the survey [HMT11]. We cite these and other related works throughout the paper and refer to [PQZb, Section 11] on further bibliography. For a natural extension of our present work, one can combine randomized matrix multiplication with randomized augmentation and additive preprocessing of [PGMQ], [PIMR10], [PQ10], [PQ12], [PQZC], [PQZb], [PY09].

1.4 Organization of the paper and selective reading

In the next section we recall some definitions and basic results. We estimate the condition numbers of Gaussian random matrices in Section 3 and of randomized matrix products in Section 4, where we also comment on numerical stabilization of GENP. In Sections 5 and 6 we apply randomized matrix multiplication to approximate the leading and trailing singular spaces of a matrix having a small numerical rank. We also approximate this matrix by a low-rank matrix, and point out applications to tensor decomposition and to approximation by structured matrices. In Section 7 we cover numerical tests, which constitute the contribution of the second author. In the Appendix we estimate the probability that a random matrix has full rank under the uniform probability distribution.

2 Some definitions and basic results

We assume computations in the field $\mathbb{R}$ of real numbers.

Hereafter "flop" stands for "arithmetic operation"; "expect" and "likely" mean "with probability 1 or close to 1" (we do not use the concept of the expected value), and the concepts "large", "small", "near", "closely approximate", "ill conditioned" and "well conditioned" are quantified in the context. Next we recall and extend some customary definitions of matrix computations [GL96], [S98].

2.1 Some basic definitions on matrix computations

$\mathbb{R}^{m\times n}$ is the class of real $m \times n$ matrices $A = (a_{i,j})_{i,j}^{m,n}$.

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

$\mathbf{e}_i$ is the ith coordinate vector of dimension n for $i = 1, \dots, n$. These vectors define the identity matrix $I_n = (\mathbf{e}_1 \mid \dots \mid \mathbf{e}_n)$ of size $n \times n$. $O_{k,l}$ is the $k \times l$ matrix filled with zeros. We write I and O where the size of a matrix is not important or is defined by context.

A^T is the transpose of a matrix A.

2.2 Range, rank, and generic rank profile

$\mathcal{R}(A)$ denotes the range of an $m \times n$ matrix A, that is the linear space $\{\mathbf{z} : \ \mathbf{z} = A\mathbf{x}\}$ generated by its columns. $\mathrm{rank}(A) = \dim \mathcal{R}(A)$ denotes its rank. $A_k^{(k)}$ denotes the leading, that is northwestern $k \times k$ block submatrix of a matrix A. A matrix of a rank ρ has *generic rank profile* if all its leading $i \times i$ blocks are nonsingular for $i = 1, \dots, \rho$. If such matrix is nonsingular itself, then it is called *strongly nonsingular.*

Fact 2.1. *The set* $\mathbb{M}$ *of* $m \times n$ *matrices of rank* ρ *is an algebraic variety of dimension* $(m+n-\rho)\rho$.

Proof. Let M be an $m \times n$ matrix of a rank ρ with a nonsingular leading $\rho \times \rho$ block M_{00} and write $M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$. Then the $(m-\rho) \times (n-\rho)$ *Schur complement* $M_{11} - M_{10}M_{00}^{-1}M_{01}$ must vanish, which imposes $(m-\rho)(n-\rho)$ algebraic equations on the entries of M. Similar argument can be applied where any $\rho \times \rho$ submatrix of the matrix M (among $\binom{m}{\rho}\binom{n}{\rho}$ such submatrices) is nonsingular. Therefore $\dim \mathbb{M} = mn - (m-\rho)(n-\rho) = (m+n-\rho)\rho$. $\square$

2.3 Orthogonal, Toeplitz and circulant matrices

A real matrix Q is called *orthogonal* if $Q^TQ = I$ or $QQ^T = I$. In Section 7 we write $Q(A)$ to denote a unique orthogonal matrix specified by the following result.

Fact 2.2. *[GL96, Theorem 5.2.2]. QR factorization $A = QR$ of a matrix A having full column rank into the product of an orthogonal matrix $Q = Q(A)$ and an upper triangular matrix $R = R(A)$ is unique provided that the factor R is a square matrix with positive diagonal entries.*

A *Toeplitz* $m \times n$ matrix $T_{m,n} = (t_{i-j})_{i,j=1}^{m,n}$ is defined by its first row $(t_{-h})_{h=0}^{n-1}$ and the subvector $(t_h)_{h=1}^{n-1}$ of its first column vector. Circulant matrices are the subclass of Toeplitz matrices where $t_g = t_h$ if $|g-h| = n$.

Theorem 2.1. *$O((m+n)\log(m+n))$ flops suffice to multiply an $m \times n$ Toeplitz matrix by a vector.*

2.4 Norms, SVD, generalized inverse, and singular spaces

$||A||_h$ is the h-norm and $||A||_F = \sqrt{\sum_{i,j=1}^{m,n}|a_{i,j}|^2}$ is the Frobenius norm of a matrix $A = (a_{i,j})_{i,j=1}^{m,n}$. We write $||A|| = ||A||_2$ and $||\mathbf{v}|| = \sqrt{\mathbf{v}^T\mathbf{v}} = ||\mathbf{v}||_2$ and recall from [GL96, Section 2.3.2 and Corollary 2.3.2] that

$$\max_{i,j=1}^{m,n}|a_{i,j}| \le ||A|| = ||A^T|| \le \sqrt{mn}\,\max_{i,j=1}^{m,n}|a_{i,j}|,$$

$$\frac{1}{\sqrt{m}}||A||_1 \le ||A|| \le \sqrt{n}||A||_1,\quad ||A||_1 = ||A^T||_\infty,\quad ||A||^2 \le ||A||_1||A||_\infty, \tag{2.1}$$

$$||A|| \le ||A||_F \le \sqrt{n}\ ||A||, \tag{2.2}$$

$$||AB||_h \le ||A||_h||B||_h \text{ for } h = 1, 2, \infty \text{ and any matrix product } AB. \tag{2.3}$$

Define an *SVD* or *full SVD* of an $m \times n$ matrix A of a rank ρ as follows,

$$A = S_A\Sigma_AT_A^T. \tag{2.4}$$

Here $S_AS_A^T = S_A^TS_A = I_m$, $T_AT_A^T = T_A^TT_A = I_n$, $\Sigma_A = \operatorname{diag}(\widehat{\Sigma}_A, O_{m-\rho,n-\rho})$, $\widehat{\Sigma}_A = \operatorname{diag}(\sigma_j(A))_{j=1}^{\rho}$, $\sigma_j = \sigma_j(A) = \sigma_j(A^T)$ is the jth largest singular value of a matrix A for $j = 1,\dots,\rho$, and we write $\sigma_j = 0$ for $j > \rho$. These values have the minimax property

$$\sigma_j = \max_{\dim(\mathbb{S})=j}\ \min_{\mathbf{x}\in\mathbb{S},\ ||\mathbf{x}||=1} ||A\mathbf{x}||,\ j = 1,\dots,\rho, \tag{2.5}$$

where $\mathbb{S}$ denotes linear spaces [GL96, Theorem 8.6.1]. Consequently $\sigma_\rho > 0$, $\sigma_1 = \max_{||\mathbf{x}||=1}||A\mathbf{x}|| = ||A||$.

Fact 2.3. *If A_0 is a submatrix of a matrix A, then $\sigma_j(A) \ge \sigma_j(A_0)$ for all j.*

Proof. [GL96, Corollary 8.6.3] implies the claimed bound where A_0 is any block of columns of the matrix A. Transposition of a matrix and permutations of its rows and columns do not change singular values, and thus we can extend the bounds to all submatrices A_0. □

$A^+ = T_A\operatorname{diag}(\widehat{\Sigma}_A^{-1}, O_{n-\rho,m-\rho})S_A^T$ is the Moore–Penrose pseudo-inverse of the matrix A of (2.4), and

$$||A^+|| = 1/\sigma_\rho(A) \tag{2.6}$$

for a matrix A of a rank ρ. A^{+T} stands for $(A^+)^T = (A^T)^+$, and A^{-T} stands for $(A^{-1})^T = (A^T)^{-1}$.

In Sections 5–6 we use the following definitions. For every integer k in the range $1 \le k < \operatorname{rank}(A)$ define the partition $S_A = (S_{k,A} \mid S_{A,m-k})$ and $T_A = (T_{k,A} \mid T_{A,n-k})$ where the submatrices $S_{k,A}$ and $T_{k,A}$ are formed by the first k columns of the matrices S_A and T_A, respectively. Write $\Sigma_{k,A} = \operatorname{diag}(\sigma_j(A))_{j=1}^k$, $\mathbb{S}_{k,A} = \mathcal{R}(S_{k,A})$ and $\mathbb{T}_{k,A} = \mathcal{R}(T_{k,A})$. If $\sigma_k > \sigma_{k+1}$, then $\mathbb{S}_{k,A}$ and $\mathbb{T}_{k,A}$ are the left and right *leading singular spaces*, respectively, associated with the k largest singular values of the matrix A, whereas their orthogonal complements $\mathbb{S}_{A,m-k} = \mathcal{R}(S_{A,m-k})$ and $\mathbb{T}_{A,n-k} = \mathcal{R}(T_{A,n-k})$

are the left and right *trailing singular spaces*, respectively, associated with the other singular values of A. The pairs of subscripts $\{k, A\}$ versus $\{A, m-k\}$ and $\{A, n-k\}$ mark the leading versus trailing singular spaces. The left singular spaces of A are the right singular spaces of A^T and vice versa. All matrix bases for the singular spaces $\mathbb{S}_{k,A}$ and $\mathbb{T}_{k,A}$ are given by matrices $S_{k,A}X$ and $T_{k,A}Y$, respectively, for nonsingular $k \times k$ matrices X and Y. Orthogonal matrices X and Y define orthogonal matrix bases for these spaces. B is an *approximate matrix basis* for a space $\mathbb{S}$ within a relative error norm bound τ if there exists a matrix E such that $B + E$ is a matrix basis for this space $\mathbb{S}$ and if $||E|| \le \tau ||B||$.

2.5 Condition number, numerical rank and generic conditioning profile

$\kappa(A) = \frac{\sigma_1(A)}{\sigma_\rho(A)} = ||A||\ ||A^+||$ is the condition number of an $m \times n$ matrix A of a rank ρ. Such matrix is *ill conditioned* if $\sigma_1(A) \gg \sigma_\rho(A)$ and is *well conditioned* otherwise. See [D83], [GL96, Sections 2.3.2, 2.3.3, 3.5.4, 12.5], [H02, Chapter 15], [KL94], [S98, Section 5.3], on the estimation of matrix norms and condition numbers.

An $m \times n$ matrix A has *numerical rank*, denoted $\mathrm{nrank}(A)$ and not exceeding $\mathrm{rank}(A)$, if the ratios $\sigma_j(A)/||A||$ are small for $j > \mathrm{nrank}(A)$ but not for $j \le \mathrm{nrank}(A)$.

Remark 2.1. *One can specify the adjective "small" above as "smaller than a fixed positive tolerance". The choice of the tolerance can be a challenge, e.g., for the matrix* $\mathrm{diag}(1.1^{-j})_{j=0}^{999}$.

If a well conditioned $m \times n$ matrix A has a rank $\rho < l = \min\{m, n\}$, then almost all its close neighbours have full rank l (see Section 3.2), and all of them have numerical rank ρ. Conversely, suppose a matrix A has a positive numerical rank $\rho = \mathrm{nrank}(A)$ and *truncate its SVD* by setting to 0 all its singular values, except for the ρ largest ones. Then the resulting matrix $A - E$ is well conditioned and has rank ρ and $||E|| = \sigma_{\rho+1}(A)$, and so $A - E$ is a rank-ρ approximation to the matrix A within the error norm bound $\sigma_{\rho+1}(A)$. At a lower computational cost we can obtain rank-ρ approximations of the matrix A from its rank-revealing factorizations [GE96], [HP92], [P00a], and we further decrease the computational cost by applying randomized algorithms in Section 5.

An $m \times n$ matrix has *generic conditioning profile* (cf. the end of Section 2.2) if it has a numerical rank ρ and if its leading $i \times i$ blocks are nonsingular and well conditioned for $i = 1, \ldots, \rho$. If such matrix has full rank (that is if $\rho = \min\{m, n\}$) and if it is well conditioned itself, then we call it *strongly well conditioned*. The following theorem shows that GENP and block Gaussian elimination applied to a strongly well conditioned matrix are numerically safe.

Theorem 2.2. *Cf. [PQZa, Theorem 5.1]. Assume GENP or block Gaussian elimination applied to an $n \times n$ matrix A and write $N = ||A||$ and $N_- = \max_{j=1}^{n} ||(A_j^{(j)})^{-1}||$. Then the absolute values of all pivot elements of GENP and the norms of all pivot blocks of block Gaussian elimination do not exceed $N + N_- N^2$, whereas the absolute values of the reciprocals of these elements and the norms of the inverses of the blocks do not exceed N_-.*

3 Ranks and conditioning of Gaussian random matrices

3.1 Random variables and Gaussian random matrices

Definition 3.1. *$F_\gamma(y) = Probability\{\gamma \le y\}$ (for a real random variable γ) is the* cumulative distribution function (cdf) *of γ evaluated at y. $F_{g(\mu,\sigma)}(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} \exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$ for a Gaussian random variable $g(\mu, \sigma)$ with a mean μ and a positive variance σ^2, and so*

$$\mu - 4\sigma \le y \le \mu + 4\sigma \text{ with a probability near 1.} \tag{3.1}$$

Definition 3.2. *A matrix (or a vector) is a* Gaussian random matrix (or vector) *with a mean μ and a positive variance σ^2 if it is filled with independent identically distributed Gaussian random variables, all having the mean μ and variance σ^2. $\mathcal{G}_{\mu,\sigma}^{m\times n}$ is the set of such Gaussian random $m \times n$ matrices, which are* standard *for $\mu = 0$ and $\sigma^2 = 1$. By restricting this set to $m \times n$ Toeplitz matrices* *where only the* $m+n-1$ *entries of the first row and column are independent we obtain the set of* $\mathcal{T}_{\mu,\sigma}^{m\times n}$ Gaussian random Toeplitz *matrices . Likewise we obtain the set* $\mathcal{Z}_{\mu,\sigma}^{n\times n}$ *of* Gaussian random circulant matrices, *where only* n *entries of the first row are independent.*

3.2 Nondegeneration of Gaussian random matrices

The total degree of a multivariate monomial is the sum of its degrees in all its variables. The total degree of a polynomial is the maximal total degree of its monomials.

Lemma 3.1. *[DL78], [S80], [Z79]. For a set* Δ *of a cardinality* $|\Delta|$ *in any fixed ring let a polynomial in* m *variables have a total degree* d *and let it not vanish identically on this set. Then the polynomial vanishes in at most* $d|\Delta|^{m-1}$ *points.*

We assume that Gaussian random variables range over infinite sets Δ, usually over the real line or its interval. Then the lemma implies that a nonzero polynomial vanishes with probability 0. Consequently a Gaussian random general, Toeplitz or circulant matrix has generic rank profile with probability 1 because the determinant of any its block is a polynomials in the entries. Likewise Gaussian random general, Toeplitz and circulant matrices have generic rank profile with probability 1. Hereafter, wherever this causes no confusion, we assume by default that *Gaussian random general, Toeplitz and circulant matrices have generic rank profile.* This property can be readily extended to the products and various functions of general, sparse and structured Gaussian random matrices. Similar properties hold with probability near 1 where the random variables are sampled under the uniform probability distribution from a finite set of a large cardinality (see the Appendix).

3.3 Extremal singular values of Gaussian random matrices

Besides having full rank with probability 1, Gaussian random matrices in Definition 3.2 are likely to be well conditioned [D88], [E88], [ES05], [CD05], [B11], and even the sum $M+A$ for $M\in\mathbb{R}^{m\times n}$ and $A\in\mathcal{G}_{\mu,\sigma}^{m\times n}$ is likely to be well conditioned unless the ratio $\sigma/||M||$ is small or large [SST06].

The following theorem states an upper bound proportional to y on the cdf $F_{1/||A^+||}(y)$, that is on the probability that the smallest positive singular value $1/||A^+||=\sigma_l(A)$ of a Gaussian random matrix A is less than a nonnegative scalar y (cf. (2.6)) and consequently on the probability that the norm $||A^+||$ exceeds a positive scalar x. The stated bound still holds if we replace the matrix A by $A-B$ for any fixed matrix B, and for $B=O_{m,n}$ the bounds can be strengthened by a factor $y^{|m-n|}$ [ES05], [CD05].

Theorem 3.1. *Suppose* $A\in\mathcal{G}_{\mu,\sigma}^{m\times n}$*,* $B\in\mathbb{R}^{m\times n}$*,* $l=\min\{m,n\}$*,* $x>0$*, and* $y\geq 0$*. Then* $F_{\sigma_l(A-B)}(y)\leq 2.35\sqrt{l}y/\sigma$*, that is Probability*$\{||(A-B)^+||\geq 2.35x\sqrt{l}/\sigma\}\leq 1/x$*.*

Proof. For $m=n$ this is [SST06, Theorem 3.3]. Apply Fact 2.3 to extend it to any pair $\{m,n\}$. □

The following two theorems supply lower bounds $F_{||A||}(z)$ and $F_{\kappa(A)}(y)$ on the probabilities that $||A||\leq z$ and $\kappa(A)\leq y$ for two scalars y and z, respectively, and a Gaussian random matrix A. We do not use the second theorem, but state it for the sake of completeness and only for square $n\times n$ matrices A. The theorems imply that the functions $1-F_{||A||}(z)$ and $1-F_{\kappa(A)}(y)$ decay as $z\to\infty$ and $y\to\infty$, respectively, and that the decays are exponential in $-z^2$ and proportional to $\sqrt{\log y}/y$, respectively. For small values $y\sigma$ and a fixed n the lower bound of Theorem 3.3 becomes negative, in which case the theorem becomes trivial. Unlike Theorem 3.1, in both theorems we assume that $\mu=0$.

Theorem 3.2. *[DS01, Theorem II.7]. Suppose* $A\in\mathcal{G}_{0,\sigma}^{m\times n}$*,* $h=\max\{m,n\}$ *and* $z\geq 2\sigma\sqrt{h}$*. Then* $F_{||A||}(z)\geq 1-\exp(-(z-2\sigma\sqrt{h})^2/(2\sigma^2))$*, and so the norm* $||A||$ *is likely to have order* $\sigma\sqrt{h}$*.*

Theorem 3.3. *[SST06, Theorem 3.1]. Suppose* $0<\sigma\leq 1$*,* $y\geq 1$*,* $A\in\mathcal{G}_{0,\sigma}^{n\times n}$*. Then the matrix* A *has full rank with probability* 1 *and* $F_{\kappa(A)}(y)\geq 1-(14.1+4.7\sqrt{(2\ln y)/n})n/(y\sigma)$*.*

Proof. See [SST06, the proof of Lemma 3.2]. □

4 Condition numbers of randomized matrix products and generic preconditioning

Next we deduce probabilistic lower bounds on the smallest singular values of the products of fixed and random matrices. We begin with three lemmas. The first of them is obvious, the second easily follows from minimax property (2.5).

Lemma 4.1. *$\sigma_j(SM) = \sigma_j(MT) = \sigma_j(M)$ for all j if S and T are square orthogonal matrices.*

Lemma 4.2. *Suppose $\Sigma = \operatorname{diag}(\sigma_i)_{i=1}^n$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, $G \in \mathbb{R}^{r\times n}$, $H \in \mathbb{R}^{n\times r}$. Then $\sigma_j(G\Sigma) \geq \sigma_j(G)\sigma_n$, $\sigma_j(\Sigma H) \geq \sigma_j(H)\sigma_n$ for all j. If also $\sigma_n > 0$, then $\operatorname{rank}(G\Sigma) = \operatorname{rank}(G)$, $\operatorname{rank}(\Sigma H) = \operatorname{rank}(H)$.*

Lemma 4.3. *[SST06, Proposition 2.2]. Suppose $H \in \mathcal{G}_{\mu,\sigma}^{m\times n}$, $SS^T = S^TS = I_m$, $TT^T = T^TT = I_n$. Then $SH \in \mathcal{G}_{\mu,\sigma}^{m\times n}$ and $HT \in \mathcal{G}_{\mu,\sigma}^{m\times n}$.*

The following theorem implies that multiplication by standard Gaussian random matrix is unlikely to decrease the smallest positive singular value of a matrix dramatically, even though $UV = O$ for some pairs of rectangular orthogonal matrices U and V.

Theorem 4.1. *Suppose $G' \in \mathcal{G}_{\mu,\sigma}^{r\times m}$, $H' \in \mathcal{G}_{\mu,\sigma}^{n\times r}$, $M \in \mathbb{R}^{m\times n}$, $G = G' + U$, $H = H' + V$ for some matrices U and V, $r(M) = \operatorname{rank}(M)$, $x > 0$ and $y \geq 0$. Then $F_{1/||(GM)^+||}(y) \leq F(y, M, \sigma)$ and $F_{1/||(MH)^+||}(y) \leq F(y, M, \sigma)$ for $F(y, M, \sigma) = 2.35y\sqrt{\widehat{r}}||M^+||/\sigma$ and $\widehat{r} = \min\{r, r(M)\}$, that is $Probability\{||P^+|| \geq 2.35x\sqrt{\widehat{r}}||M^+||/\sigma\} \leq 1/x$ for $P = GM$ and $P = MH$.*

Proof. With probability 1, the matrix MH has rank $\widehat{r}$ because $H \in \mathcal{G}_{\mu,\sigma}^{n\times r}$. So (cf. (2.6))

$$F_{1/||(MH)^+||}(y) = F_{\sigma_{\widehat{r}}(MH)}(y). \tag{4.1}$$

Let $M = S_M\Sigma_M T_M^T$ be full SVD where $\Sigma_M = \operatorname{diag}(\widehat{\Sigma}_M, O) = \Sigma_M \operatorname{diag}(I_{r(M)}, O)$ and $\widehat{\Sigma}_M = \operatorname{diag}(\sigma_j(M))_{j=1}^{r(M)}$ is a nonsingular diagonal matrix. We have $MH = S_M\Sigma_M T_M^T H$, and so $\sigma_j(MH) = \sigma_j(\Sigma_M T_M^T H)$ for all j by virtue of Lemma 4.1, because S_M is a square orthogonal matrix. Write $H_{r(M)} = (I_{r(M)} \mid O)T_M^T H$ and observe that $\sigma_j(\Sigma_M T_M^T H) = \sigma_j(\widehat{\Sigma}_M H_{r(M)})$ and consequently

$$\sigma_j(MH) = \sigma_j(\widehat{\Sigma}_M H_{r(M)}) \text{ for all } j. \tag{4.2}$$

Combine equation (4.2) for $j = \widehat{r}$ with Lemma 4.2 for the pair (Σ, H) replaced by $(\widehat{\Sigma}_M, H_{r(M)})$ and obtain that $\sigma_{\widehat{r}}(MH) \geq \sigma_{r(M)}(M)\sigma_{\widehat{r}}(H_{r(M)}) = \sigma_{\widehat{r}}(H_{r(M)})/||M^+||$. We have $T_M^T H' \in \mathcal{G}_{\mu,\sigma}^{n\times r}$ by virtue of Lemma 4.3, because T_M is a square orthogonal matrix; consequently $H_{r(M)} = H'_{r(M)} + B$ for $H'_{r(M)} \in \mathcal{G}_{\mu,\sigma}^{r(M)\times r}$ and some matrix B. Therefore we can apply Theorem 3.1 for $A = H'_{r(M)}$ and obtain the bound of Theorem 4.1 on $F_{1/||(MH)^+||}(y)$. One can similarly deduce the bound on $F_{1/||(GM)^+||}(y)$ or can just apply the above bound on $F_{1/||(MH)^+||}(y))$ for $H = G^T$ and M replaced by M^T and then recall that $(M^TG^T)^T = GM$. ◻

By combining (2.3) with Theorems 3.2 (for $B = O$) and 4.1 we can probabilistically bound the condition numbers of randomized products GM and MH. The following corollary extends the bound of Theorem 4.1 for a randomized matrix product to the bounds for its blocks.

Corollary 4.1. *Suppose j, k, m, n, q and s are integers, $1 \leq j \leq q$, $1 \leq k \leq s$, $M \in \mathbb{R}^{m\times n}$, $\sigma > 0$, $G \in \mathcal{G}_{\mu,\sigma}^{q\times m}$, $H \in \mathcal{G}_{\mu,\sigma}^{n\times s}$, $\operatorname{rank}(M_j) = j$ for $M_j = M\begin{pmatrix} I_j \\ O_{n-j,j}\end{pmatrix}$, $\operatorname{rank}(M^{(k)}) = k$ for $M^{(k)} = (I_k \mid O_{k,m-k})M$, and $y \geq 0$. Then (i) with probability 1 the matrix GM (resp. MH) has full rank if $\operatorname{rank}(M) \geq q$ (resp. if $\operatorname{rank}(M) \geq s$). Furthermore (ii) $F_{1/||((GM)_j^{(j)})^+||}(y) \leq 2.35y\sqrt{j}||M_j^+||/\sigma$ if $\operatorname{rank}(M) \geq j$, $F_{1/||((MH)_k^{(k)})^+||}(y) \leq 2.35y\sqrt{k}||(M^{(k)})^+||/\sigma$ if $\operatorname{rank}(M) \geq k$.*

Proof. We immediately verify part (i) by applying the techniques of Section 3.2. To prove part (ii) apply Theorem 4.1 replacing G by $(I_j \mid O_{j,q-j})G$ and replacing M by $M\begin{pmatrix} I_j \\ O_{n-j,j} \end{pmatrix}$. For every k apply Theorem 4.1 replacing M by $(I_k \mid O_{k,m-k})M$ and replacing H by $H\begin{pmatrix} I_k \\ O_{s-k,k} \end{pmatrix}$. □

Corollary 4.1 can be immediately extended to any block of the matrices GM and MH, but we single out the leading blocks because applications of GENP and block Gaussian elimination are numerically safe where these blocks are nonsingular and well conditioned. We have empirical evidence that such applications are numerically safe even where we use circulant multipliers G and H filled with ± 1 and where randomization is restricted to choosing the signs $\pm$ (see Tables 7.1 and 7.3). The study in [T11] provides some partial formal support for somewhat similar empirical observations.

5 Approximate bases for singular spaces, low-rank approximation, and the computation of numerical rank

5.1 Randomized low-rank approximation: an outline and an extension to approximation by structured matrices

Supppose we seek a rank-ρ approximation to a matrix A that has a numerical rank ρ. We can solve this problem by computing the SVD of the matrix A or its rank-revealing factorization [GE96], [HP92], [P00a], but in this section we study alternative numerically stable and noncostly solution based on randomized matrix multiplication. As by-product we obtain approximate matrix bases for the left or right leading singular space $\mathbb{T}_{\rho,A}$ and $\mathbb{S}_{\rho,A}$.

Let us supply further details. Our next theorem expresses a rank-ρ approximation to a matrix A through an approximate matrix basis for the (left or right) leading singular space $\mathbb{T}_{\rho,A}$ or $\mathbb{S}_{\rho,A}$. If we are given a reasonably small upper bound ρ_+ on ρ, then with a probability near 1 we can readily obtain such a basis for the space $\mathbb{T}_{\rho,A}$ from the product A^TG for $G \in \mathcal{G}_{0,1}^{m\times\rho_+}$. Theorem 5.2 of Section 5.3 formally supports correctness of this randomized algorithm, but our tests support it consistently even where $G \in \mathcal{T}_{0,1}^{m\times\rho_+}$ (see Tables 7.2 and 7.3), and we conjecture that the same is true for various other classes of sparse and structured multipliers G defined by much fewer random parameters compared to the number of the entries. We specify a low-rank approximation algorithm in Section 5.4, which has important applications to matrix computations, many listed in [HMT11].

Here is a simple sample extension. Assume a matrix W having a possibly unknown numerical displacement rank d, that is lying near some matrices with a small displacement rank d (see the definitions in [KKM79], [BM01], [P01]). We can compute one of these displacements as a rank-d approximation to the displacement of the matrix W, and then immediately recover a structured matrix approximating the matrix W.

5.2 Low-rank approximation via the basis of a leading singular space

The following theorem expresses a rank-q approximation (within an error norm $\sigma_{q+1}(A)$) to a matrix A through a matrix basis of its leading singular space $\mathbb{T}_{q,A}$ or $\mathbb{S}_{q,A}$.

Theorem 5.1. *Suppose A is an $m\times n$ matrix, $S_A\Sigma_AT_A^T$ is its SVD of (2.4), q is a positive integer, $q \le \min\{m,n\}$, and T and S are matrix bases for the spaces $\mathbb{T}_{q,A}$ and $\mathbb{S}_{q,A}$, respectively. Then*

$$||A - AT(T^TT)^{-1}T^T|| = ||A - S(S^TS)^{-1}S^TA|| = \sigma_{q+1}(A). \tag{5.1}$$

For orthogonal matrices T and S we have $T^TT = S^TS = I_q$ and

$$||A - ATT^T|| = ||A - SS^TA|| = \sigma_{q+1}(A). \tag{5.2}$$

Proof. Let us first write $P = T_{q,A}T_{q,A}^T$ and $r = n - q$ and estimate the norm $||A - AP||$. We have $AP = S_A\Sigma_A T_A^T T_{q,A}T_{q,A}^T$. Substitute $T_A^T T_{q,A} = \begin{pmatrix} I_q \\ O_{r,q} \end{pmatrix}$ and obtain $AP = S_A\Sigma_A \begin{pmatrix} T_{q,A}^T \\ O_{r,q} \end{pmatrix}$, whereas $A = S_A\Sigma_A \begin{pmatrix} T_{q,A}^T \\ T_{A,r}^T \end{pmatrix}$. Therefore

$$A - AP = S_A\Sigma_A \begin{pmatrix} O_{q,n} \\ T_{A,r}^T \end{pmatrix} = S_A \operatorname{diag}(O_q, \operatorname{diag}(\sigma_j)_{j=q+1}^n) \begin{pmatrix} O_{q,n} \\ T_{A,r}^T \end{pmatrix},$$

and so $||A - AP|| = ||\operatorname{diag}(\sigma_j)_{j=q+1}^n|| = \sigma_{q+1}$ because S_A and $T_{A,r}$ are orthogonal matrices. Similarly deduce that $||A - S_{q,A}S_{q,A}^T A|| = \sigma_{q+1}(A)$. This proves (5.1) and (5.2) for $T = T_{q,A}$ and $S = S_{q,A}$.

Now let the matrices T and S have full rank, $\mathcal{R}(T) = \mathbb{T}_{q,A} = \mathcal{R}(T_{q,A})$, $\mathcal{R}(S) = \mathbb{S}_{q,A} = \mathcal{R}(S_{q,A})$, and so $T = T_{q,A}U$ and $S = S_{q,A}V$ for two nonsingular matrices U and V. Consequently $T(T^TT)^{-1}T^T = T_{q,A}U(U^TT_{q,A}^TT_{q,A}U)^{-1}U^TT_{q,A}^T$. Substitute $T_{q,A}^TT_{q,A} = I_q$ and deduce that $(U^TT_{q,A}^TT_{q,A}U)^{-1} = (U^TU)^{-1} = U^{-1}U^{-T}$. Therefore $U(U^TT_{q,A}^TT_{q,A}U)^{-1}U^T = UU^{-1}U^{-T}U^T = I_q$, and so $T(T^TT)^{-1}T^T = T_{q,A}U(U^TT_{q,A}^TT_{q,A}U)^{-1}U^TT_{q,A}^T = T_{q,A}T_{q,A}^T$. Similarly $S(S^TS)^{-1}S^T = S_{q,A}S_{q,A}^T$, implying the desired extension. □

5.3 A basis of a leading singular space via randomized products

The following theorem supports randomized approximation of matrix bases for the leading singular spaces $\mathbb{T}_{\rho,A}$ and $\mathbb{S}_{\rho,A}$ of a matrix A having numerical rank ρ.

Theorem 5.2. *Suppose a matrix $A \in \mathbb{R}^{m\times n}$ has a numerical rank ρ, $H \in \mathcal{G}_{0,1}^{n\times\rho_+}$ and $G \in \mathcal{G}_{0,1}^{m\times\rho_+}$ for $\rho_+ \geq \rho$. Then the matrices $T = A^TG$ and $S = AH$ have full rank with probability 1 and we can expect that they have numerical rank ρ and that*

$$S + \Delta = S_{\rho,A}U \text{ and } T + \Delta' = T_{\rho,A}V \tag{5.3}$$

for two matrices Δ and Δ' having norms of order $\sigma_{\rho+1}(A)$ and for two nonsingular matrices U and V having condition numbers of at most order $||A||/(\sigma_\rho(A)\sqrt{\rho})$.

Proof. The techniques of Section 3.2 and Theorem 4.1 support the claims about ranks and numerical ranks. It remains to deduce the former probabilistic relationship $\mathbb{S}_{\rho,AH+\Delta} = \mathbb{S}_{\rho,A}$ of (5.3) because we can apply it to A^T to obtain the latter relationship $\mathbb{T}_{\rho,A^TG+\Delta'} = \mathbb{T}_{\rho,A}$.

Assume the SVD $A = S_A\Sigma_AT_A^T$ and note that $||\Sigma_A - \operatorname{diag}(\Sigma_{\rho,A}, O_{m-\rho,n-\rho})|| \leq \sigma_{\rho+1}(A)$. Consequently $||A - S_A \ \operatorname{diag}(\Sigma_{\rho,A}, O_{m-\rho,n-\rho})\ T_A^T|| \leq \sigma_{\rho+1}(A)$ and $AH = S - \Delta$, $S = S_{\rho,A}U$, $||\Delta|| \leq \sigma_{\rho+1}(A)\ ||AH||$ where $U = \Sigma_{\rho,A}B$, $B = T_{\rho,A}^TH$, and we can expect that the norm $||H||$ is bounded from above and below by two positive constants (see Theorem 3.2). This implies (5.3). It remains to estimate $\kappa(U)$.

With probability 1 the $\rho\times\rho$ matrices B and U are nonsingular (see Section 3.2). Furthermore we have $||U|| \leq ||\Sigma_{\rho,A}||\ ||B||$ where $||\Sigma_{\rho,A}|| = ||A||$ and $||B|| \leq ||T_{\rho,A}||\ ||H|| = ||H||$. So $||U|| \leq ||A||\ ||H|| = O(||A||)$. We also have $||U^+|| \leq ||\Sigma_{\rho,A}^{-1}||\ ||B^{-1}||$ for nonsingular matrix B. Observe that $||\Sigma_{\rho,A}^{-1}|| = 1/\sigma_\rho(A)$, apply Theorem 4.1 where $M = T_{\rho,A}^T$, $\widehat{r} = \rho$ and $\sigma_{r(M)}(M) = \sigma = 1$ and obtain that the norm $||B^{-1}||$ is likely to have at most order $1/\sqrt{\rho}$. Summarizing we can expect that the norm $||U^+||$ has at most order $1/(\sigma_\rho(A)\sqrt{\rho})$. Consequently $\kappa(U) = ||U||\ ||U^+||$ has at most order $||A||/(\sigma_\rho(A)\sqrt{\rho})$. □

Remark 5.1. *The theorem suggests using multiplication by random matrices for approximation of the leading singular spaces $\mathbb{T}_{\rho,A}$ and $\mathbb{S}_{\rho,A}$ of a matrix A given with its numerical rank ρ. The approximation is facilitated as the gaps increase between the singular values of the input matrix A. This motivates using the power transforms $A \Longrightarrow B_h = (AA^T)^hA$ for positive integers h because $\sigma_j(B_h) = (\sigma_j(A))^{2h+1}$ for all j.*

Remark 5.2. *Two matrix bases of the trailing singular space $\mathbb{T}_{A,n-\rho}$ and of the leading singular space $\mathbb{T}_{\rho,A}$ of an $m \times n$ matrix A are orthogonal to one another, and similarly for any pair of matrix bases of the spaces $\mathbb{S}_{A,m-\rho}$ and $\mathbb{S}_{\rho,a}$. This duality can help simplify the computation and approximation of the bases.*

5.4 A prototype algorithm for low-rank approximation

Together Theorems 5.1 and 5.2 imply correctness of the following prototype algorithm where we assume that the input matrix has an unknown numerical rank and we know its upper bound. The algorithm employs approximation of a leading singular space of the input matrix.

Proto-Algorithm 5.1. Rank-ρ approximation of a matrix *(cf. [HMT11, Section 10.3]).*

INPUT: *A matrix $A \in \mathbb{R}^{m\times n}$ having an unknown numerical rank ρ, an integer $\rho_+ \geq \rho$, and two tolerances τ and τ' of order $\sigma_{\rho+1}(A)/||A||$. (We can choose τ at Stage 2 based on rank revealing factorization of an auxiliary $n \times \rho_+$ matrix. The computation of this factorization is noncostly where ρ is small. We can choose τ' at Stage 3 based on the required output accuracy, and can adjust both tolerances if the algorithm fails to produce a satisfactory output.)*

OUTPUT: *FAILURE (with a low probability) or an integer ρ and two matrices $T \in \mathbb{R}^{n\times\rho}$ and $A_\rho \in \mathbb{R}^{m\times n}$, both having ranks at most ρ and such that $||A_\rho - A|| \leq \tau'||A||$ and T satisfies bound (5.3) of Theorem 5.2 for $||\Delta'|| \leq \tau||A||$.*

COMPUTATIONS:

1. *Compute the $n \times \rho_+$ matrix $T' = A^TG$ for $G \in \mathcal{G}_{0,1}^{m\times\rho_+}$.*
2. *Compute a rank revealing factorization of the matrix T' and choose the minimal integer s and an $n \times s$ matrix T such that $||T' - (T \mid O_{n,\rho_+-s})|| \leq \tau||A||$.*
3. *Compute the matrix $A_s = AT(T^TT)^{-1}T^T$. Output $\rho = s$, T and A_ρ and stop if $||A_\rho - A|| \leq \tau'||(A)||$. Otherwise output FAILURE and stop.*

Assume a proper choice of both tolerances τ and τ'. Then by virtue of Theorem 5.2, we can expect that at Stage 2 we obtain $s = \rho$ and an approximate matrix basis T for the singular space $\mathbb{T}_{\rho,A}$ (within an error norm of at most order $\sigma_{\rho+1}(A)$). If so, Stage 3 outputs FAILURE with a probability near 0, by virtue of Theorems 5.1, and in the case of FAILURE we can reapply the algorithm for new values of random parameters or for the adjusted tolerance values τ and τ'. At Stage 2 we have $s \leq \rho$ because $\text{nrank}(A^TG) \leq \text{nrank}(A) = \rho$, whereas for a sufficiently small tolerance τ' the bound $||A_\rho - A|| \leq \tau'||(A)||$ at Stage 3 implies that $s \geq \text{nrank}(A)$. These observations enable us to certify correctness of the outputs ρ, T, and A_ρ of the algorithm.

We can similarly approximate the matrix A by a rank-ρ matrix $S(S^TS)^{-1}S^TA$, by first computing the matrix $S' = AH$ for $H \in \mathcal{G}_{0,1}^{n\times\rho_+}$, then computing its rank revealing factorization, which is expected to define an approximate matrix basis S for the space $\mathbb{S}_{\rho,A}$, and finally applying Theorem 5.1, to approximate the matrix A by a rank-ρ matrix.

Remark 5.3. *By applying rank revealing QR factorization at Stage 2 of the algorithm we could produce an orthogonal matrix T and consequently simplify Stage 3 by computing $A_s = ATT^T$ (cf. (5.2)). We adopted such a variation of the algorithm in our tests in Section 7.*

Remark 5.4. *For larger integer q we can substantially simplify Stage 1 of the algorithm by choosing structured multipliers G from the class of the* subsample random Fourier transforms *or* SRFTs. *Under this choice the estimated probability of obtaining low rank approximation is close to the case of Gaussian random multipliers G. Our tests in Section 7 provide informal empirical support for similar use of random Toeplitz multipliers G.*

Remark 5.5. *One can weaken reliability of the output to simplify Stage 3 by testing whether $||K^T(A - A_\rho)L|| \leq \tau||K||\ ||A||\ ||L||$ for matrices $K \in \mathcal{G}_{0,1}^{m\times\rho'}$ and $L \in \mathcal{G}_{0,1}^{n\times\rho''}$ and for two small positive integers ρ' and ρ'', possibly for $\rho' = \rho'' = 1$, instead of testing whether $||A_\rho - A|| \leq \tau'||(A)||$. One can similarly simplify Stage 2.*

Remark 5.6. *For $\rho_+ = \rho$ Stage 2 can be omitted because the matrix A^TG is expected to be a desired approximate matrix basis by virtue of Theorem 5.2. The increase of the dimension ρ_+ beyond ρ (called* oversampling *in [HMT11]) is relatively inexpensive if the bound ρ_+ is small. [HMT11] suggests using small oversampling even if the numerical rank ρ is known, because we have*

$$\text{Probability } \{||A - ATT^T|| \le (1 + 9\sqrt{\rho_+ \min\{m,n\}})\sigma_{\rho+1}(A)\} \ge 1 - 3(\rho_+ - \rho)^{\rho-\rho_+} \text{ for } \rho_+ > \rho.$$

Theorem 5.2, however, bounds the norm $||A - ATT^T||$ strongly also for $\rho = \rho_+$, in good accordance with the data of Tables 7.2 and 7.3. Similar results have been obtained in [T11] in the case of multipliers G defining SRFTs.

6 Application to Tensor Train decomposition

Let

$$\mathbf{A} = [A(i_1, \ldots, i_d)] \tag{6.1}$$

denote a d-dimensional *tensor* with entries $A(i_1, \ldots, i_d)$ and *spacial indices* $i_1, \ldots, i_d$ ranging from 1 to $n_1, \ldots, n_d$, respectively. Define the $d-1$ *unfolding matrices* $A_k = [A(i_1 \ldots i_k; i_{k+1} \ldots i_d)]$, $k = 1, \ldots, d$, where the semicolon separates the multi-indices $i_1 \ldots i_k$ and $i_{k+1} \ldots i_d$, which define the rows and columns of the matrix A_k, respectively, $k = 1, \ldots, d$. The paper [O09] proposed the following class of *Tensor Train Decompositions*, hereafter referred to as *TT Decompositions*, where the *summation indices* $\alpha_1, \ldots, \alpha_{d-1}$ ranged from 1 to *compression ranks* $r_1, \ldots, r_{d-1}$, respectively,

$$T = \sum_{\alpha_1, \ldots, \alpha_{d-1}} G_1(i_1, \alpha_1) G_2(\alpha_1, i_1, \alpha_2) \cdots G_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) G_d(\alpha_d, i_d). \tag{6.2}$$

Theorem 6.1. *[O09]. For any tensor $\mathbf{A}$ of (6.1) there exists a TT decomposition (6.2) such that $\mathbf{A} = \mathbf{T}$ and $r_k = \text{rank}(A_k)$ for $k = 1, \ldots, d-1$.*

There is a large and growing number of important applications of TT decompositions (6.2) to modern computations (cf. e.g., [OT09], [OT10], [OT11]) where the numerical ranks of the unfolding matrices A_k are much smaller than their ranks, and it is desired to compress TT decompositions respectively.

Theorem 6.2. *[OT10]. For any tensor $\mathbf{A}$ of (6.1) and any set of positive integers $r_k \le \text{rank}(A_k)$, $k = 1, \ldots, d-1$, there exists a TT decomposition (6.2) such that*

$$||\mathbf{A} - \mathbf{T}||_F^2 \le \sum_{k=1}^{d-1} \tau_k^2,\ \tau_k = \min_{\text{rank}(B) = r_k} ||A_k - B||_F,\ k = 1, \ldots, d-1. \tag{6.3}$$

The constructive proof of this theorem in [OT10] relies on inductive approximation of unfolding matrices by their SVDs truncated to the compression ranks r_k. Let us sketch this construction. For $d = 2$ we obtain a desired TT decomposition $T(i_1, i_2) = \sum_{\alpha_1}^{r_1} G_1(i_1, \alpha_1) G_2(\alpha_1, i_2)$ (that is a sum of r_1 outer products of r_1 pairs of vectors) simply by truncating the SVD of the matrix $A(i_1, i_2)$. At the inductive step one truncates the SVD of the first unfolding matrix $A_1 = S_{A_1} \Sigma_{A_1} T_{A_1}^T$ to obtain rank-r_1 approximation of this matrix $B_1 = S_{B_1} \Sigma_{B_1} T_{B_1}^T$ where $\Sigma_{B_1} = \text{diag}(\sigma_j(A_1))_{j=1}^{r_1}$ and the matrices S_{B_1} and T_{B_1} are formed by the first r_1 columns of the matrices S_{A_1} and T_{A_1}, respectively. Then it remains to approximate the tensor $\mathbf{B} = [B(i_1, \ldots, i_d)]$ represented by the matrix B_1. Rewrite it as $\sum_{\alpha_1=1}^{n} S_{B_1}(i_1; \alpha_1) \widehat{A}(\alpha_1; i_2 \ldots i_d)$ for $\widehat{A} = \sum_{B_1} T_{B_1}^T$, represent $\widehat{A}$ as the tensor $\widehat{\mathbf{A}} = [A(\alpha_1 i_2, i_3, \ldots, i_d)]$ of dimension $d-1$, apply the inductive hypothesis to obtain a TT-approximation of this tensor, and extend it to a TT-approximation of the original tensor $\mathbf{A}$.

In [OT10] the authors specify this construction as their Algorithm 1, prove error norm bound (6.3), then point out that the "computation of the truncated SVD for large scale and possibly dense unfolding matrices ... is unaffordable in many dimensions", propose "to replace SVD by some other dyadic decompositions $A_k \approx UV^T$, which can be computed with low complexity", and finally specify such recipe as [OT10, Algorithm 2], which is an iterative algorithm for skeleton or pseudoskeleton decomposition of matrices and which they use at Stages 5 and 6 of their Algorithm 1. The cost of each iteration of [OT10, Algorithm 2] is quite low, and empirically the iteration converges fast, but the authors welcome alternative recipes having formal support.

Proto-Algorithm 5.1 can serve as an alternative to [OT10, Algorithm 2]. For the input matrix A_1 above we use $O(r_1)$ multiplications of this matrix by $O(r_1)$ vectors, which means a low computational cost for sparse and structured inputs, whereas the expected output is an approximate matrix basis for the space $\mathbb{S}_{r_1,A_1}$ or $\mathbb{T}_{r_1,A_1}$ and a rank-r_1 approximation to the matrix A_1, within an expected error norm in $O(\sigma_{r_1+1}(A_1))$. This is the same order as in [OT10, Algorithm 1], but now we do not use SVDs. One can further decrease the error bound by means of small oversampling of Remark 5.6 and the power transform of Remark 5.1.

Remark 6.1. *A huge bibliography on tensor decompositions and on thier application to fundamental matrix computations has been recently surveyed in [KB09], but with the omission of the early works [P72], [P79], [B80], [P84], [B86], where nontrivial tensor decompositions helped to accelerate the fundamental operation of matrix multiplication, probably the first application of this kind.*

7 Numerical Experiments

Our numerical experiments with random general, Hankel, Toeplitz and circulant matrices have been performed in the Graduate Center of the City University of New York on a Dell server with a dual core 1.86 GHz Xeon processor and 2G memory running Windows Server 2003 R2. The test Fortran code has been compiled with the GNU gfortran compiler within the Cygwin environment. Random numbers have been generated with the random_number intrinsic Fortran function, assuming the uniform probability distribution over the range $\{x : -1 \le x < 1\}$. The tests have been designed by the first author and performed by his coauthor.

7.1 GENP with random circulant multipliers

Table 7.1 shows the results of our tests of the solution of a nonsingular well conditioned linear system $A\mathbf{y} = \mathbf{b}$ of n equations whose coefficient matrix has ill conditioned $n/2 \times n/2$ leading principal block for $n = 64, 256, 1024$. We have performed 100 numerical tests for each dimension n and computed the maximum, minimum and average relative residual norms $||A\mathbf{y} - \mathbf{b}||/||\mathbf{b}||$ as well as standard deviation. GENP applied to these systems outputs corrupted solutions with residual norms ranging from 10 to 10^8. When we preprocessed the systems with circulant multipliers filled with ± 1 (choosing the n signs $\pm$ at random), the norms decreased to at worst 10^{-7} for all inputs. Table 7.1 also shows further decrease of the norm in a single step of iterative refinement. Table 2 in [PQZa] shows similar results of the tests where the input matrices have been chosen similarly but so that their every leading $k \times k$ block had numerical rank k or $k-1$ and where Householder multipliers $I_n - \mathbf{u}\mathbf{v}^T/\mathbf{u}^T\mathbf{v}$ replaced the circulant multipliers. Here $\mathbf{u}$ and $\mathbf{v}$ denote two vectors filled with integers 1 and -1 under random choice of the signs $+$ and $-$.

7.2 Approximation of the tails and heads of SVDs and low-rank approximation of a matrix

At some specified stages of our tests of this subsection we performed additions, subtractions and multiplications with infinite precision (hereafter referred to as *error-free ring operations*). At the other stages we performed computations with double precision, and we rounded to double precision all random values. We performed at most two refinement iterations for the computed solution of every linear system of equations and matrix inverse.

Tables 7.2 and 7.3 display the data from our tests on the approximation of leading singular spaces of the SVD of an $n \times n$ matrix A having numerical rank q and on the approximation of this matrix with a matrix of rank ρ. For $n = 64, 128, 256$ and $\rho = 1, 8, 32$ we generated $n \times n$ random orthogonal matrices S and T and diagonal matrices $\Sigma = \mathrm{diag}(\sigma_j)_{j=1}^n$ such that $\sigma_j = 1/j,\ j = 1, \ldots, \rho,$

$\sigma_j = 10^{-10}$, $j = \rho + 1, \ldots, n$ (cf. [H02, Section 28.3]). Then we applied error-free ring operations to compute the input matrices $A = S_A \Sigma_A T_A^T$, for which $||A|| = 1$ and $\kappa(A) = 10^{10}$. Furthermore we generated random $n \times \rho$ matrices G (for $\rho = 1, 8, 32$) and successively computed the matrices $B_{\rho,A} = A^T G$, $T_{\rho,A}$, $B_{\rho,A} Y_{\rho,A}$ as a least-squares approximation to $T_{\rho,A}$, $Q_{\rho,A} = Q(B_{\rho,A})$ (cf. Fact 2.2), and $A - AQ_{\rho,A}(Q_{\rho,A})^T$ (by applying error-free ring operations). Table 7.2 summarizes the data on the residual norms $\mathrm{rn}^{(1)} = ||B_{\rho,A} Y_{\rho,A} - T_{\rho,A}||$ and $\mathrm{rn}^{(2)} = ||A - AQ_{\rho,A}(Q_{\rho,A})^T||$ obtained in 100 runs of our tests for every pair of n and ρ.

We have also performed similar tests where we generated random Toeplitz $n \times \rho$ matrices T (for $\rho = 8, 32$) and then replaced the above approximate matrix bases $B_{\rho,A} = A^T G$ for the leading singular space $\mathbb{T}_{\rho,A}$ by the matrices $B_{\rho,A} = A^T T$. Table 7.3 displays the results of these tests. In both Tables 7.2 and 7.3 the residual norms are more or less equally small.

Table 7.1: Relative residual norms: randomized circulant GENP for well conditioned linear systems with ill conditioned leading blocks (cf. [PQZa, Table 2])

dimension	**iterations**	**min**	**max**	**mean**	**std**
64	0	4.7×10^{-14}	8.0×10^{-11}	4.0×10^{-12}	1.1×10^{-11}
64	1	1.9×10^{-15}	5.3×10^{-13}	2.3×10^{-14}	5.4×10^{-14}
256	0	1.7×10^{-12}	1.4×10^{-7}	2.0×10^{-9}	1.5×10^{-8}
256	1	8.3×10^{-15}	4.3×10^{-10}	4.5×10^{-12}	4.3×10^{-11}
1024	0	1.7×10^{-10}	4.4×10^{-9}	1.4×10^{-9}	2.1×10^{-9}
1024	1	3.4×10^{-14}	9.9×10^{-14}	6.8×10^{-14}	2.7×10^{-14}

Table 7.2: Heads of SVDs and low-rank approximation by using random multipliers G

q	rrn_i	n	**min**	**max**	**mean**	**std**
1	$\mathrm{rn}^{(1)}$	64	2.35×10^{-10}	1.32×10^{-07}	3.58×10^{-09}	1.37×10^{-08}
1	$\mathrm{rn}^{(1)}$	128	4.41×10^{-10}	3.28×10^{-08}	3.55×10^{-09}	5.71×10^{-09}
1	$\mathrm{rn}^{(1)}$	256	6.98×10^{-10}	5.57×10^{-08}	5.47×10^{-09}	8.63×10^{-09}
1	$\mathrm{rn}^{(2)}$	64	8.28×10^{-10}	1.32×10^{-07}	3.86×10^{-09}	1.36×10^{-08}
1	$\mathrm{rn}^{(2)}$	128	1.21×10^{-09}	3.28×10^{-08}	3.91×10^{-09}	5.57×10^{-09}
1	$\mathrm{rn}^{(2)}$	256	1.74×10^{-09}	5.58×10^{-08}	5.96×10^{-09}	8.47×10^{-09}
8	$\mathrm{rn}^{(1)}$	128	2.56×10^{-09}	1.16×10^{-06}	4.30×10^{-08}	1.45×10^{-07}
8	$\mathrm{rn}^{(1)}$	256	4.45×10^{-09}	3.32×10^{-07}	3.40×10^{-08}	5.11×10^{-08}
8	$\mathrm{rn}^{(2)}$	64	1.46×10^{-09}	9.56×10^{-08}	5.77×10^{-09}	1.06×10^{-08}
8	$\mathrm{rn}^{(2)}$	128	1.64×10^{-09}	4.32×10^{-07}	1.86×10^{-08}	5.97×10^{-08}
8	$\mathrm{rn}^{(2)}$	256	2.50×10^{-09}	1.56×10^{-07}	1.59×10^{-08}	2.47×10^{-08}
32	$\mathrm{rn}^{(1)}$	64	6.80×10^{-09}	2.83×10^{-06}	1.01×10^{-07}	3.73×10^{-07}
32	$\mathrm{rn}^{(1)}$	128	1.25×10^{-08}	6.77×10^{-06}	1.28×10^{-07}	6.76×10^{-07}
32	$\mathrm{rn}^{(1)}$	256	1.85×10^{-08}	1.12×10^{-06}	1.02×10^{-07}	1.54×10^{-07}
32	$\mathrm{rn}^{(2)}$	64	1.84×10^{-09}	6.50×10^{-07}	2.30×10^{-08}	8.28×10^{-08}
32	$\mathrm{rn}^{(2)}$	128	3.11×10^{-09}	1.45×10^{-06}	2.87×10^{-08}	1.45×10^{-07}
32	$\mathrm{rn}^{(2)}$	256	4.39×10^{-09}	2.16×10^{-07}	2.37×10^{-08}	3.34×10^{-08}

8 Conclusions

It is well known that random matrices tend to be well conditioned, and this property motivates our application of random matrix multipliers for advancing some fundamental matrix computations. We

Table 7.3: Heads of SVDs and low-rank approximations by using random Toeplitz multipliers T

q	$\mathrm{rrn}^{(i)}$	n	**min**	**max**	**mean**	**std**
8	$\mathrm{rrn}^{(1)}$	64	2.22×10^{-09}	7.89×10^{-06}	1.43×10^{-07}	9.17×10^{-07}
8	$\mathrm{rrn}^{(1)}$	128	3.79×10^{-09}	4.39×10^{-05}	4.87×10^{-07}	4.39×10^{-06}
8	$\mathrm{rrn}^{(1)}$	256	5.33×10^{-09}	3.06×10^{-06}	6.65×10^{-08}	3.12×10^{-07}
8	$\mathrm{rrn}^{(2)}$	64	1.13×10^{-09}	3.66×10^{-06}	6.37×10^{-08}	4.11×10^{-07}
8	$\mathrm{rrn}^{(2)}$	128	1.81×10^{-09}	1.67×10^{-05}	1.90×10^{-07}	1.67×10^{-06}
8	$\mathrm{rrn}^{(2)}$	256	2.96×10^{-09}	1.25×10^{-06}	2.92×10^{-08}	1.28×10^{-07}
32	$\mathrm{rrn}^{(1)}$	64	6.22×10^{-09}	5.00×10^{-07}	4.06×10^{-08}	6.04×10^{-08}
32	$\mathrm{rrn}^{(1)}$	128	2.73×10^{-08}	4.88×10^{-06}	2.57×10^{-07}	8.16×10^{-07}
32	$\mathrm{rrn}^{(1)}$	256	1.78×10^{-08}	1.25×10^{-06}	1.18×10^{-07}	2.03×10^{-07}
32	$\mathrm{rrn}^{(2)}$	64	1.64×10^{-09}	1.26×10^{-07}	9.66×10^{-09}	1.48×10^{-08}
32	$\mathrm{rrn}^{(2)}$	128	5.71×10^{-09}	9.90×10^{-07}	5.50×10^{-08}	1.68×10^{-07}
32	$\mathrm{rrn}^{(2)}$	256	4.02×10^{-09}	2.85×10^{-07}	2.74×10^{-08}	4.48×10^{-08}

first prove the basic fact that with a probability close to 1 multiplication by a Gaussian random matrix does not increase the condition number of a matrix and of its any block dramatically compared to the condition number of the input matrix. As an immediate implication random multipliers are likely to stabilize numerically GENP (that is Gaussian elimination with no pivoting) and block Gaussian elimination applied to a nonsingular and well conditioned matrix, possibly having ill conditioned and singular leading blocks. by applying to input matrix randomized structured multipliers. Another basic fact states that with a probability close to 1 the column sets of the products A^TG and AH where an $m \times n$ matrix A has a numerical rank ρ and G and H are Gaussian random matrices of sizes $m \times \rho$ and $n \times \rho$, respectively, approximate some bases for the left and right leading singular spaces $\mathbb{S}_{\rho,A}$ and $\mathbb{T}_{\rho,A}$ associated with the ρ largest singular values of the matrix A. Having any of such approximate bases available we can readily approximate the matrix A by a matrix of rank ρ, This has further well known extensions to many important matrix computations, and we point out two new ones, to the approximation of a matrix by a structured matrix lying nearby and to Tensor Train decomposition. Finding new extensions and applications is a natural subject for further study.

Our extensive tests consistently show efficiency of the proposed techniques even where instead of general Gaussian random multipliers we apply structured and sparse multipliers where randomization is limited to much fewer random parameters or just to the choice of the signs $\pm$ of a few auxiliary vectors. The recent paper [T11] is an important step toward understanding and exploiting this phenomenon and should motivate further research effort. Another natural research subject is the combination of randomized matrix multiplication with randomized techniques of additive preprocessing and augmentation, recently studied in [PGMQ], [PIMR10], [PQ10], [PQ12], [PQZC], [PQZa], [PQZb], and [PQZc].

Appendix

A Uniform random sampling and nonsingularity of random matrices

Uniform random sampling of elements from a finite set Δ is their selection from this set at random, independently of each other and under the uniform probability distribution on the set Δ.

Theorem A.1. *Under the assumptions of Lemma 3.1 let the values of the variables of the polynomial be randomly and uniformly sampled from a finite set Δ. Then the polynomial vanishes with a probability at most $\frac{d}{|\Delta|}$.*

Corollary A.1. *Let the entries of a general or Toeplitz $m \times n$ matrix have been randomly and uniformly sampled from a finite set Δ of cardinality $|\Delta|$ (in any fixed ring). Let $l = \min\{m, n\}$. Then (a) every $k \times k$ submatrix M for $k \le l$ is nonsingular with a probability at least $1 - \frac{k}{|\Delta|}$ and (b) is strongly nonsingular with a probability at least $1 - \sum_{i=1}^{k} \frac{i}{|\Delta|} = 1 - \frac{(k+1)k}{2|\Delta|}$.*

Proof. The claimed properties of nonsingularity and nonvanishing hold for generic matrices. The singularity of a $k \times k$ matrix means that its determinant vanishes, but the determinant is a polynomial of total degree k in the entries. Therefore Theorem A.1 implies parts (a) and consequently (b). Part (c) follows because a fixed entry of the inverse vanishes if and only if the respective entry of the adjoint vanishes, but up to the sign the latter entry is the determinant of a $(k-1) \times (k-1)$ submatrix of the input matrix M, and so it is a polynomial of degree $k - 1$ in its entries. □

Acknowledgements: Our research has been supported by NSF Grant CCF–1116736 and PSC CUNY Awards 64512–0042 and 65792–0043. We are also grateful to Mr. Jesse Wolf for helpful comments.

References

[A94] O. Axelsson, *Iterative Solution Methods*, Cambridge Univ. Press, England, 1994.

[B80] D. Bini, Border Rank of $p \times q \times 2$ Tensors and the Optimal Approximation of a Pair of Bilinear Forms, in *Lecture Notes in Computer Science*, **85**, 98–108, Springer, 1980.

[B86] D. Bini, Border Rank of $m \times n \times (mn - q)$ Tensors, *Linear Algebra and Its Applications*, **79**, 45–51, 1986.

[B02] M. Benzi, Preconditioning Techniques for Large Linear Systems: a Survey, *J. of Computational Physics*, **182**, 418–477, 2002.

[B11] C. Beltrán, Estimates on the Condition Number of Random, Rank-deficient Matrices, *IMA Journal of Numerical Analysis*, **31, 1**, 25–39, 2011.

[BG05] A. Böttcher, S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*, SIAM Publications, Philadelphia, 2005.

[BM01] D. A. Bini, B. Meini, Approximate Displacement Rank and Applications, in *AMS Conference "Structured Matrices in Operator Theory, Control, Signal and Image Processing"*, Boulder, 1999 (edited by V. Olshevsky), *American Math. Society*, 215–232, Providence, RI, 2001.

[CD05] Z. Chen, J. J. Dongarra, Condition Numbers of Gaussian Random Matrices, *SIAM. J. on Matrix Analysis and Applications*, **27**, 603–620, 2005.

[D83] J. D. Dixon, Estimating Extremal Eigenvalues and Condition Numbers of Matrices, *SIAM J. on Numerical Analysis*, **20, 4**, 812–814, 1983.

[D88] J. Demmel, The Probability That a Numerical Analysis Problem Is Difficult, *Math. of Computation*, **50**, 449–480, 1988.

[DL78] R. A. Demillo, R. J. Lipton, A Probabilistic Remark on Algebraic Program Testing, *Information Processing Letters*, **7**, **4**, 193–195, 1978.

[DS01] K. R. Davidson, S. J. Szarek, Local Operator Theory, Random Matrices, and Banach Spaces, in *Handbook on the Geometry of Banach Spaces* (W. B. Johnson and J. Lindenstrauss editors), pages 317–368, North Holland, Amsterdam, 2001.

[E88] A. Edelman, Eigenvalues and Condition Numbers of Random Matrices, *SIAM J. on Matrix Analysis and Applications*, **9**, **4**, 543–560, 1988.

[ES05] A. Edelman, B. D. Sutton, Tails of Condition Number Distributions, *SIAM J. on Matrix Analysis and Applications*, **27**, **2**, 547–560, 2005.

[G97] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.

[GE96] M. Gu, S. C. Eisenstat, Efficient algorithms for computing a strong rank-revealing QR factorization, *SIAM Journal on Scientific Computing*, **17**, 848–869, 1996.

[GL96] G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 1996 (third addition).

[H02] N. J. Higham, *Accuracy and Stability in Numerical Analysis*, SIAM, Philadelphia, 2002 (second edition).

[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, The rank revealing QR decomposition and SVD, *Math. of Computation*, **58**, 213–232, 1992.

[KB09] T. G. Kolda, B. W. Bader, Tensor Decompositions and Applications, *SIAM Review*, **51, 3**, 455–500, 2009.

[KKM79] T. Kailath, S. Y. Kung, M. Morf, Displacement Ranks of Matrices and Linear Equations, *Journal Math. Analysis and Appls*, **68, 2**, 395–407, 1979.

[KL94] C. S. Kenney, A. J. Laub, Small-Sample Statistical Condition Estimates for General Matrix Functions, *SIAM J. on Scientific and Statistical Computing*, **15**, 36–61, 1994.

[O09] I. V. Oseledets, A Compact Matrix Form of the d-Dimensional Tensor Decomposition, Preprint 2009-01, *INM RAS*, March 2009.

[OT09] I. V. Oseledets, E. E. Tyrtyshnikov, Breaking the Curse of Dimensionality, or How to Use SVD in Many Dimensions, *SIAM J. Scientific Comp.*, **31**, **5**, 3744–3759, 2009.

[OT10] I. V. Oseledets, E. E. Tyrtyshnikov, TT-cross Approximation for Multidimensional Arrays, *Linear Algebra Appls.* **432, 1**, 70–88, 2010.

[OT11] I. Oseledets, E. E. Tyrtyshnikov, Algebraic Wavelet Transform via Quantics Tensor Train Decomposition, *SIAM J. Sci. Comp.*, **33**, **3**, 1315–1328, 2011.

[P72] V. Y. Pan, On Schemes for the Evaluation of Products and Inverses of Matrices (in Russian), *Uspekhi Matematicheskikh Nauk*, **27**, **5 (167)**, 249–250, 1972.

[P79] V. Y. Pan, Fields Extension and Trilinear Aggregating, Uniting and Canceling for the Acceleration of Matrix Multiplication, *Proceedings of the 20th Annual IEEE Symposium on Foundations of Computer Science (FOCS 79)*, 28–38, IEEE Computer Society Press, Long Beach, California, 1979.

[P84] V. Y. Pan, How Can We Speed up Matrix Multiplication? *SIAM Review,* **26**, **3**, 393–415, 1984.

[P00a] C.–T. Pan, On the Existence and Computation of Rank-revealing LU Factorization, *Linear Algebra and Its Applications*, **316**, 199–222, 2000.

[P01] V. Y. Pan, *Structured Matrices and Polynomials: Unified Superfast Algorithms*, Birkhäuser/Springer, Boston/New York, 2001.

[PGMQ] V. Y. Pan, D. Grady, B. Murphy, G. Qian, R. E. Rosholt, A. Ruslanov, Schur Aggregation for Linear Systems and Determinants, *Theoretical Computer Science*, *Special Issue on Symbolic–Numerical Algorithms* (D. A. Bini, V. Y. Pan, and J. Verschelde editors), **409**, **2**, 255–268, 2008.

[PIMR10] V. Y. Pan, D. Ivolgin, B. Murphy, R. E. Rosholt, Y. Tang, X. Yan, Additive Preconditioning for Matrix Computations, *Linear Algebra and Its Applications*, **432**, 1070–1089, 2010.

[PQ10] V. Y. Pan, G. Qian, Randomized Preprocessing of Homogeneous Linear Systems of Equations, *Linear Algebra and Its Applications*, **432**, 3272–3318, 2010.

[PQ12] V. Y. Pan, G. Qian, Solving Linear Systems of Equations with Randomization, Augmentation and Aggregation, *Linear Algebra and Its Applications*, **437**, 2851–1876, 2012.

[PQa] V. Y. Pan, G. Qian, Condition Numbers of Random Toeplitz and Circulant Matrices, Tech. Report TR 2012013, *PhD Program in Comp. Sci.*, *Graduate Center, CUNY*, 2012.

Available at http://www.cs.gc.cuny.edu/tr/techreport.php?id=442

[PQZa] V. Y. Pan, G. Qian, A. Zheng, Randomized Preprocessing versus Pivoting, *Linear Algebra and Its Applications*, in print. http://dx.doi.org/10.1016/j.laa.2011.02.052

[PQZb] V. Y. Pan, G. Qian, A. Zheng, Randomized Matrix Computations, Tech. Report TR 2012009, *PhD Program in Comp. Sci.*, *Graduate Center, CUNY*

Available at http://www.cs.gc.cuny.edu/tr/techreport.php?id=438 and

http://arxiv.org/abs/1210.7476

[PQZc] V. Y. Pan, G. Qian, A. Zheng, Randomized Augmentation and Additive Preprocessing, Tech. Report TR 201201x, *PhD Program in Comp. Sci.*, *Graduate Center, CUNY* 2012.

[PQZC] V. Y. Pan, G. Qian, A. Zheng, Z. Chen, Matrix Computations and Polynomial Rootfinding with Preprocessing, *Linear Algebra and Its Applications*, **434**, 854–879, 2011.

[PY09] V. Y. Pan, X. Yan, Additive Preconditioning, Eigenspaces, and the Inverse Iteration, *Linear Algebra and Its Applications*, **430**, 186–203, 2009.

[S80] J. T. Schwartz, Fast Probabilistic Algorithms for Verification of Polynomial Identities, *Journal of ACM*, **27**, **4**, 701–717, 1980.

[S98] G. W. Stewart, *Matrix Algorithms, Vol I: Basic Decompositions*, SIAM, 1998.

[SST06] A. Sankar, D. Spielman, S.-H. Teng, Smoothed Analysis of the Condition Numbers and Growth Factors of Matrices, *SIAM J. on Matrix Analysis*, **28**, **2**, 446–476, 2006.

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

[Z79] R. E. Zippel, Probabilistic Algorithms for Sparse Polynomials, *Proceedings of EUROSAM'79, Lecture Notes in Computer Science*, **72**, 216–226, Springer, Berlin, 1979.